Sensing Contacts, Coughs, and Hand Hygiene

William Huang

Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/EECS-2020-212
http://www2.eecs.berkeley.edu/Pubs/TechRpts/2020/EECS-2020-212.html

December 17, 2020

Copyright © 2020, by the author(s).
All rights reserved.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission.

Sensing Contacts, Coughs, and Hand Hygiene

by

William Huang

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Computer Science

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Associate Professor Prabal K. Dutta, Chair
Professor David E. Culler
Professor John Chuang

Fall 2020

Sensing Contacts, Coughs, and Hand Hygiene

Copyright 2020
by
William Huang

Abstract

Sensing Contacts, Coughs, and Hand Hygiene

by

William Huang

Doctor of Philosophy in Computer Science

University of California, Berkeley

Associate Professor Prabal K. Dutta, Chair

Detailed contact tracing that not only captures the social interaction graph, but also precise interaction distance and duration could prove useful in a wide variety of applications. Most notably, we have seen this play out in the global COVID-19 pandemic, where social distancing and contact tracing have proven critical in efforts to combat disease spread. Traditionally, contact tracing has relied on manual reporting, which provides only coarse grained data and relies on the subjectivity of human memory. These factors have led to a drive for wearable sensor based solutions which can provide objective face-to-face interaction data. Ideally, these sensors would provide precise interaction distances and durations, and would only report these metrics when users are actually facing each other and are not separated by a barrier. Current contact tracing sensors can generally be divided into two camps. First are sensors that can provide precise interaction distances, but require infrastructure to run, making them difficult to deploy in practice. Second are sensors that do not require infrastructure, but only provide a rough sense of proximity, making it difficult to analyze which interactions are significant. The majority of these systems also cannot determine if there is a barrier separating users, or if the users are facing each other.

To address these issues, we present Opo, a wearable sensor which requires no infrastructure to run, provides interaction distance accurate to 5 cm, and only records interaction distances when users are facing each other with no barriers between them. The key problem we identify is that systems that provide precise interaction distances require RF based neighbor discovery protocols to synchronize nodes before performing ranging operations to get interaction distance. Instead, Opo utilizes ultrasonic passive vigilance, to perform neighbor discovery and ranging at the same time, lowering system complexity and power usage.

In addition, while current wearables for contact tracing have largely focused on detecting interactions, in practice this information is greatly enriched by knowledge of health behaviors and symptoms. For example, researchers are often interested in detecting hand-washing behavior due to its importance in combating a wide variety of infectious diseases. Current

hand washing sensor systems generally use wrist mounted accelerometers or utilize smart badges and soap-dispenser mounted sensors. Of these two system types, only smart badge plus soap dispenser sensor systems are able to capture if a person uses soap, a key consideration when measuring hand washing behavior. However, smart badge systems only detect when a person washes their hands with soap, and do not sense hand washing duration. The key problem is that current smart badge systems use low-resolution ranging technologies, making it difficult for them to determine when a user approaches leaves a sink. To address this problem, we create a smart badge plus dispenser mounted sensor system by extending Opo with passive vigilance in the accelerometric domain. This extension allows soap dispenser mounted Opos to passively detect when a dispenser is used and provide precise times when a user approaches and leaves a sink. To the best of our knowledge, our hand washing system is the first that can detect and categorize both soaped and un-soaped hand washing events and measure hand washing duration.

Researchers are also often interested in when people first experience symptom onsets. In particular, researchers are often interested in when people begin coughing, due to its prominence as an early symptom in many infectious diseases. Current cough sensing systems focus on counting the number of times a person coughs over a given period of time. These sensors require a user to wear a voice recorder and record all of their audio over the period of time. These systems then identify and count coughs in post-processing. This technique has shown very promising results, but requires a massive invasion of user privacy, making them difficult to deploy in many situations. In addition, our review of prior work on coughing shows that in many applications, simply knowing when a person starts coughing or general trends in a person's cough counts provides significant value. To fill this niche, we create CoughNote, a wearable privacy preserving cough sensor. Instead of constantly recording audio, CoughNote utilizes passive vigilance in the audio domain to capture 1 s snippets of potential coughs, while avoiding recording sensitive vocalized audio such as speech. These potential coughs can then be analyzed in post processing without violating user privacy. Although CoughNote does not capture every cough, it can show general cough trends while preserving usability and being smaller, lighter, and almost three times as long lived as a typical voice recorder.

Overall, our work creates a wearable sensing kit that researchers can use to study face-to-face interactions and important contextual health information. We have conducted two pilot studies using Opos and CoughNote with epidemiologists, and hope that our sensors enable future gains in better understanding and forecasting disease spread. Furthermore, our work shows the power of using passive vigilance to create complex, high-resolution wearables, and we hope that future wearable sensor designers draw inspiration from our designs.

To my family and friends.

Contents

Contents ii

List of Figures iv

List of Tables xi

1 Introduction 1
- 1.1 Wearables for Face-to-Face Interaction Sensing . . . 2
- 1.2 Sensing Health Context . . . 3
- 1.3 Active Sensing Limits Usability . . . 3
- 1.4 Thesis Statement . . . 4
- 1.5 Contributions of this Dissertation . . . 4

2 Tracking Face-to-Face Interactions 6
- 2.1 Motivation and Background . . . 6
- 2.2 Related Work . . . 8
- 2.3 Design Requirements and Targets . . . 12
- 2.4 Design . . . 13
- 2.5 Implementation . . . 20
- 2.6 Evaluation . . . 20
- 2.7 Reducing Angular Offset Errors . . . 29
- 2.8 Summary . . . 31

3 Deployment Experiences and Evaluation 32
- 3.1 Motivation and Background . . . 32
- 3.2 Opo vs BLE Proximity Sensing . . . 32
- 3.3 Study Design: Midwest Office . . . 33
- 3.4 Office Study Evaluation . . . 34
- 3.5 Exploring Chest Worn Sensors . . . 40
- 3.6 Study Design: Motion Capture . . . 40
- 3.7 Preliminary Results: Motion Capture . . . 43
- 3.8 Summary . . . 46

4 Tracking Hand Washing Rates **47**
4.1 Motivation and Background 47
4.2 Related Work 49
4.3 System Design 50
4.4 Evaluation 53
4.5 Summary 59

5 Privacy Preserving Cough Sensing **61**
5.1 Motivation and Background 61
5.2 Related Work 62
5.3 Design 66
5.4 Implementation 73
5.5 Evaluation 75
5.6 Summary 81

6 Conclusion **82**

Bibliography **83**

List of Figures

2.1 A survey of op-amps comparing power draw and gain. Cascading op-amps results in exponential growth in gain with only a linear growth in power. The shaded gray region contains the gain to power tradeoff from one to five op amps (left-top to right-bottom) normalized against the circled op amp we chose. 14
2.2 The Opo ultrasonic subsystem is composed of a low-power receive frontend (top) and a simple inverter-driven transmit frontend (bottom). An analog switch controls whether the transducer is transmitting or receiving. 15
2.3 An Opo ranging operation begins with a wakeup UL pulse. Upon receipt, a receiver disables its UL RX frontend and starts a timer and waits for 48 ms for any UL multipath signals to dissipate. The receiver then enables its radio in anticipation of the TDoA signals and starts a 8 ms range-failure timeout. 55 ms after the wakeup pulse, the transmitter sends a simultaneous RF packet and ranging UL pulse. Upon receipt of the RF packet, the receiver re-enables its UL RX frontend and receives the ranging UL pulse. The receiver computes the TDoA between the RF and ultrasonic pulses to calculate its range from the transmitter. If the receiver's 8 ms range-failure timer triggers before receiving the complete range sequence, it abandons the range operation and returns to idle. 17
2.4 An optimized version of Opo's ranging protocol that requires more low-level radio control. The second ultrasonic pulse is eliminated, calculating the TDoA instead between the wakeup pulse and the subsequent RF packet. The delay for sending the RF ranging packet is also reduced (40 ms to 10 ms). A 10 ms ultrasonic wakeup window covers approximately 3 m of ranging. Sensors further away would miss the radio packet as they would wake up too late to receive it. 18

2.5 Implementations and wearing scenarios of Opo. Figure 2.5a shows various implementations of Opo. The leftmost board is a standalone Opo frontend, designed to act as a daughter-board for standard mote platforms. The center board is the evaluated Opo system. The round board is optimized for form-factor, and can be worn in a case or by sticking the transducer though a button hole. Figure 2.5b shows various ways to wear Opo. The Opo in the upper left acts as a shirt button; the Opo in the upper right is a lapel pin; the Opo in the bottom left attaches to a tie clip; the Opo in bottom right is a badge clip. Finally, Figure 2.5c shows the latest iteration of Opo, which measures 1.3 x 1 in. This version has a BLE radio to enable smartphone communication 19

2.6 Controlled Microbenchmarks: Our test set up (2.6a) is comprised of Opo sensors mounted on two metal poles, a laser range finder, and a laser angle finder. 2.6b shows that our average error is usually < 2 cm. The black bars are the 95% error bars for Opo's range estimations, showing that Opo is both accurate and precise. To test angular offset performance (people chatting in a circle), we pointed two Opo sensors at the same focal point and slowly increased the angle between them, as shown in 2.6c. Our results (2.6d) show that while error does increase with angular offset, average error is only 6% at 60°(six people in a circle). 20

2.7 Four-Way Ranging Accuracy: Setup (2.7a), ground truth (2.7b), and Opo measurements (2.7c). The measurements shown in 2.7c are the median measurements between each pair of sensors, aggregated over the duration of the experiment (12 min). 21

2.8 Various Pairwise Human Interactions: 2.8a shows a typical hallway encounter where two participants walk towards each other in a hallway, stop and chat for 65 s, and then continue on their way. To test Opo's ability to capture short interactions (2.8b), two participants stood on either side of a hallway corner. At regular intervals, one participant would turn the corner, face the other for 5 s, then go back to his original side of the corner. As 2.8b shows, Opo successfully captures the majority of short interactions. 2.8c shows an extended, professional interaction where two participants sat and collaborated on a project for 75 min. The variations in distance are the result of natural movements, and show what proximity sensing systems miss. 23

2.9 Speed Dating: 3 tables are laid out, with 1 stationary participant sitting at each table, and 3 participants who rotate between tables (2.9a). Rotation times decrease from roughly 120 to 60 to 45 s. 2.9b, 2.9c, 2.9d represent the interactions at each table, with each rotating participant being represented by a different color and marker. Ground truth is roughly set to 1.5 m, but due to chair position and posture differences, exact error is difficult to determine. However, ranging is precise within interactions, and only 7 of 396 range measurements are unrealistic. Distance variations between interactions are the result of people getting in and out of table chairs. 24

2.10 Group Chat: 6 participants stand in a circle and chat for around 6.5 min. The setup and rough ground truth is shown in 9a, although participants naturally moved and shifted during the course of the experiment. In 9b, chords represent interactions between participants, with chord width being based on measured interaction time. 9c-9h show the range measurements from each participant/sensor. Rough ground truth is shown as solid lines. 14 of the 2683 measurements are found to be unrealistic. 26
2.11 Current Draw Traces: Red lines represent various components (UL TX frontend, radio, mcu, flash) waking up and performing tasks. In 2.11a, the high current draw time between 53-62 s is due to a radio driver inefficiency. The ranging reception shown in 2.11b is done at 6 m to increase the visibility of the high powered, radio reception state. The SFD graphs for 2.11a and 2.11b show when the relevant radio transmission and reception actually occur. The smaller current spikes occurring before 50 s in 2.11a and 2.11b are from the UL frontends and MCU waking up, while the larger current spikes are from the radio turning on. 27
2.12 Real-World Deployment: 2.12a shows Opo measurements from two participants during a party. Measurements from one participant's sensor is consistently validated by measurements from the other participant's sensor. 2.12b shows sample data from one participant. The shaded areas are time spent by the participant in uninstrumentable spaces (e.g. restaurants). 29
2.13 Effects of Angle of Arrival. 30
2.14 Exploring Ultrasonic Rise Time 30

3.1 Contact networks for office site A. Each node represents a participant, and a connection between two nodes indicates that those two participants had at least one interaction using that sensor. Both sensors produce similarly connected graphs. However, further examination of the data indicates that this is somewhat coincidental. The BLE sensor misses many short-duration interactions which are critical to connectivity in the Opo network, but makes up for this by falsely detecting participants in adjacent offices and cubicles as interacting. 35
3.2 The importance of short duration interactions. Here we show the Opo generated contact network with all interactions and with interactions less than 20 s filtered out. Filtering out short interactions only results has a 16% reduction in total interaction time, but results in a significantly less connected contact network. . 37
3.3 Participants ranked by normalized total interaction time and mean interaction distance. Total interaction time is only weakly correlated with mean interaction distance, with a Spearman's rank correlation coefficient of 0.16. 38

3.4 Calculating droplet exposure scores for participants. Figure 3.4a shows a simplified relationship between between interaction distance and droplet exposure [67]. We use this to calculate how many droplets each participant exposed other participants to during this study, and show normalized scores in Figure 3.4b. We find that droplet scores only correlate moderately (0.30 rank correlation) with interaction time, meaning that high-resolution interaction distance may provide significant value in better understanding disease spread. 39

3.5 Experimental setup. Figure 3.5a shows the overall layout of our work-space and where our participants sat. The desk number corresponds to the participant ID number used in our charts below. The blue cones indicate the general location of the Optitrack cameras, which were mounted to the ceiling. Figure 3.5b shows how we mounted motion capture markers to our participants. 41

3.6 Calculating relative angle. Relative angle measures how many degrees a person would have to turn to be directly facing another person. Here we see the the interaction distance d, the relative angle from person B to person A, θ_B, and the relative angle from person A to person B, θ_A. While the interaction distance between persons A and B is symmetrical, the relative angles between them need not be. E.g, if person A has to turn a shorter distance to face person B than person B has to turn to face person A, then the relative angle of A to B (θ_A) will be smaller than the relative angle of B to A (θ_B). An extreme example of this would be if person A was standing behind person B in line. Since person A is facing person B, θ_A would be zero degrees. In contrast, person B would have to turn all the way around to face person A with a θ_B of 180 degrees. If two people are are both directly facing each other, each of them has a relative angle of zero degrees to one another. 42

3.7 Face-to-Face interaction times for participant pairs. Two-sided interactions refers to interactions where both participants are facing each other, while one sided interactions refers to interactions where at least one participant is facing the other, but the other participant may be looking in a different direction. 44

3.8 Comparing body-to-body and face-to-face interaction times and distances. Body-to-body based interaction distance and time is generated from Opo sensors, while face-to-face data is from the Optitrack motion capture system. Figure 3.8b shows interactions in which both participants face each other, while Figure 3.8c show interactions in which at least one participant faces the other. Overall, Opo slightly underestimates face-to-face interaction distance. This is expected because the Optitrack system measures the distance between the center of each participant's head rather than the front of their chests. 45

4.1 System overview. The person shown is wearing an Opo sensor, and an Opo sensor is mounted to the soap dispenser. The person faces as the soap dispenser as they walk closer to wash their hands (a). When they get to the sink to wash their hands, the Opo sensors detect an extremely close distance between the person and the sink (b). In addition, when the person uses the dispenser, it triggers an accelerometer on the dispenser mounted Opo. Our system determines that this person is washing their hands due to their close proximity to the soap dispenser when it is used. When the person is done washing their hands, they walk away (c), resulting in either longer distance measurements between the person and dispenser Opos, or a lack of distance measurements between them. This lets the system know that the person is done washing their hands, and the total hand washing duration is estimated as the time between the soap dispenser trigger and the last close range measurement between the two Opos. 51

4.2 Detecting soap dispenser usage using an accelerometric trigger. Opo's accelerometer has two phases, inactive and active. When the soap dispenser is not in use, the accelerometer reads a negligible acceleration (a), and is inactive. When the accelerometer is used, the accelerometer experiences a spike in acceleration (b), causing the acceleration to exceed the programmed threshold (0.032 g). This causes the accelerometer to switch to the active state and trigger an interrupt to the Opo's mcu, letting the mcu know that the dispenser was used. After the acceleration falls below the threshold for 50 ms (c), the accelerometer assumes the person is done using the dispenser, and reverts back to the inactive state (d). 52

4.3 Comparing the effect of different time buffers on the number of sensed hand washing events (**a**) and mean sensed hand washing duration (**b**). (**C**) shows an example hand washing trace. The blue dots are Opo ranging measurements between the person's and the dispenser's Opos, and are the distance measurements recorded during a 12 s hand washing event. Because our method of cleaning the data homogenizes the number of events regardless of the selected time buffer, (**a**) is generated from data that has not been time-filtered to show the effect of changing the time buffer. 54

4.4 Histograms of Opo hand-washing data and prior work. **a** is a histogram of all hand-washing instances captured by Opo, while **b** shows a histogram of mean hand-washing data per participant in our study. **c** shows the distribution of prior work. The histogram bins are the ones used in Borchgrevink et al's paper [83]. The 15-20 s bin contains all hand-washing durations above 15 s. The 0 s hand-washing duration bin represents people who did not wash their hands. 56

4.5 Data distribution and breakdown for our study. The bins used in Borchgrevink et al's study present a misleading representation of our data, so we present more accurate histograms in **a** and **b**. IN addition, **c** shows a breakdown of all our participants. As we can see, participants 21 and 22 have a significantly higher mean hand-washing time than other participants. Based on the length of some of their hand-washing instances (40+ s), we speculate that they were doing things at the sink other than washing their hands. 57
4.6 Opo hand washing durations, with and without soap. Soaped vs un-soaped hand-washing rates lines up very well with past secret observer studies. Hand-washing durations largely match our intuition that people who do not use soap generally just quickly wet their hands. 58

5.1 5.1a shows CoughNote in its case being worn as a necklace. The case measures 1.5 x 1.6 x 0.6 in, and the case, battery, and pcb weigh a combined 15.9 g. Figure 5.1b shows our prototype implementation of CoughNote, which measures 1.3 x 1.3 in. The pcb weighs 6.7 g. 66
5.2 System Diagram. The overall architecture of CoughNote is simple. Using a COTS microphone, we stream audio to an always-on low-power (14 uA) hardware trigger. The hardware trigger activates near the beginning of coughs or similar sounds, but crucially does not activate for speech. Once activated, the mcu wakes up, a simple circuit (fast storage layer) enables the mcu to sample and save a 1 s audio clip at faster speeds than it can support with an adc alone to a small temporary external memory bank. Once done recording, the mcu fragments the audio clip based on the mcu's available RAM and transfers the fragments to the permanent storage layer, which in our current implementation is a microSD card for convenience. . 67
5.3 5.3a shows an example cough from a healthy, non-smoking individual. This cough shows all three cough phases: explosive (1), intermediate (2), and voiced (3, optional). 5.3b shows relative amplitudes and spectrograms of conversational voice (1,3), loud, high-pitched yelling (2), and a series of coughs (4). Harmonics from a normal conversational voice can reach 10 kHz, while high-pitched yelling can reach up to 15 kHz. Even high-pitched yelling does not present significant energy at 20 kHz, while the coughs contain significant energy up to 30 kHz. 5.3c shows a door shutting, which generates a high amplitude, wide-band sound, which can trigger a false positive recording. 69

5.4 Figure 5.4a shows us an audio waveform of a yell followed by a cough. Figure 5.4b shows us the approximate effect of our filter (Figure 5.4c) on Figure 5.4a. The unfiltered yell is higher amplitude than the unfiltered cough, but the filtered cough is over 3x the amplitude of the filtered yell. Figure 5.4c shows the simulated frequency response of the filter used in our cough trigger. Our filter is theoretically centered at 20 kHz, and empirically centered at 18.6 kHz. The bright blue curve shows the attenuation of our filter, while the faded curve behind it is the phase response of the filter. For this application, the phase response is relatively unimportant. 72
5.5 Fast storage layer consisting of two digital switches, an ADC, and an FRAM chip. When set to A, the switches allow the mcu to directly communicate with the FRAM chip. When set to B, the switches short the ADC's data output line to the FRAM's data input line, and shorts the two clock lines together. The MCU first uses the A setting to initialize the FRAM chip. It then uses the B setting to short both clock lines together, and by running the clock line directly stores the ADC output to the FRAM chip. 74
5.6 Current trace for recording a 1 s audio clip 75
5.7 A typical cough waveform from our wearable device. In this particular recording, two short coughs are captured. 77
5.8 Self-reported cough estimates vs CoughNote recorded coughs. In 5.8a, we count the number of audio-clips that a participant's CoughNote recorded that contains at least one cough. In 5.8b, we count the total number of coughs captured in CoughNote's recorded clips. Because participants only estimated a "number of coughs", their self-reported count is the same for both metrics. Participant cough counts are not based on detailed logs or diaries, but just represent the participants' best guess as to how much they coughed in a day. In addition, while users estimated their total cough counts for a day, they often did not necessarily wear CoughNote for the whole day. Participants wore CoughNote for a varying number of hours and a varying number of days. These numbers are aggregated over the entire time users wore CoughNote. These results suggest that CoughNote can successfully capture cough count trends, especially when we take into account multiple coughs being present in a single audio snippet. 78
5.9 Example cough waveforms from challenge study participants with recorded coughs. Subjectively, we found that cough sounds varied between coughs of an individual, and especially between coughs from different individuals. These examples illustrate the diverse array of coughs recorded by CoughNote. 79
5.10 Average cough power spectral densities (PSD) for the pilot and challenge studies. The first 100 ms of a cough roughly represents the explosive phase of the cough, while the first 500 ms represents the total duration of the first cough in an audio clip. Although there are sometimes follow-up coughs or other noise in the last 500 ms of the audio, we found that virtually all of the energy is contained in the first 500 ms of the audio clip. In both the pilot study 80

List of Tables

2.1 Comparison of Opo to systems representing various ranging techniques. Opo was the first infrastructure-free system that could character interactions at a sub-meter level. To the best of our knowledge, it is still the only one that can do so while taking into account if people are facing each other and ensuring that people are not separated by walls. 8

3.1 Office Site A Demographics 33

3.2 Comparison of centrality and interaction time between Opo and BLE generated networks. As mentioned in Section 3.4, the similar centralities, or measures of how connected a network is, between Opo and BLE is somewhat coincidental. Although we do not have ground truth for the interaction times, based on manually examining the data, our observations during the study, and anecdotal discussions with participants, suggest the Opo interaction times are more accurate. The BLE interaction times are most likely the result of participants in adjacent work spaces being falsely detected as interacting. 36

4.1 Comparison of mean hand-washing duration and soap use rates between our data and prior work [83]. Overall, we find that our soap use rates are remarkably consistent with prior work. However, our mean-duration is higher, even once we remove two outlier participants with suspiciously high mean hand-washing durations. 59

5.1 Averages percentage of total audio energy found in high-frequency bands from categories in Google Audioset and the Urban Sound 8k DataSet. We selected AudioSet categories that we believe are common in everyday life, while the Urban Sound 8k DataSet is specifically tailed to sounds that commonly cause noise complaints in urban areas. 70

Acknowledgments

First and foremost, I would like to thank my advisor, Prof. Prabal Dutta. Thank you for taking a chance on mentoring someone who had a degree in Foreign Service and had never touched a micro-processor before grad school. Your guidance has been immeasurably invaluable, and I hope you continue to inspire and mentor grad students for decades to come. I would also like to thank my dissertation committee members, Prof. David Culler and Prof. John Chuang for their support and feedback, and Prof Ruzena Bajcsy for serving on my quals committee. To Prof. Marisa Eisenberg, Prof. Allison Aiello, Paul Zivich, and Ali Walsh, thank you for your repeated collaborations and guidance in grounding my work. Without your epidemiological expertise, I would never have been able to take this path.

Funding for my research was provided by several sources, including: the Terraswarm Research Center, which was supported by the STARnet phase of the Focus Center Research Program (FCRP), a Semiconductor Research Corporation program sponsored by MARCO and DARPA; the National Science Foundation under grants CNS0964120, CNS-1111541, and CNS-1350967; the CONIX Research Center, which is funded by JUMP, a Semiconductor Research Corporation (SRC) program sponsored by DARPA; and finally, generous gifts from Intel Texas Instruments.

To Lab11 members old and new, thank you for the PhD adventure of a lifetime. During my time in grad school, I never encountered another lab with the culture and camaraderie of our lab. Without all of you, my PhD journey would have been far less enjoyable and enriching. Meghan Clark, a connoisseur who significantly increased my tea-related expenditures. Neal Jackson, a mountain biker who sometimes dabbles in engineering. Jean-Luc Watson, who once suggested our security policy should be "safety-off". Ye-Sheng Kuo, my lab neighbor for many years who taught me so much about hardware. Pat Pannuto, who once cooked a fish in a fish in a fish. Sam DeBruin, who once got me lost in the woods of Montana. Brad Campbell, the only other engineer who wanted to go to Michigan Basketball games. Branden Ghena, expert in all things citrus and board games related. Noah Klugman, lab hotelier. Ben Kempke, German Park picnic enthusiast. Josh Adkins, Shishir Patel, Rohit Ramesh, and Thomas Zachariah, have also all played a part in making Lab11 as awesome and intellectually diverse as it was during my time here.

To my friends outside of lab, you have helped me immeasurably in life. To Tianren Qi and Brent Hillenbrand, who have stuck by my side since high-school. To Ben Lipsius and Mark Fontana, randomly meeting you guys was one of the best parts of Ann Arbor. And to Reilly Hart, Katie Noethe, Sandy Moss, Jessie Chiang, Kevin Zusy, Brian Gallagher, Jung Min Woo, and the rest of NS4, thanks for all the years of friendship. I hope we all have a lifetime of adventures together.

Finally, I wish to thank my family for everything. To my parents, for working hard to make sure I had the opportunities they did not have growing up, and for supporting me in all my academic wanderings as I figured out what I wanted to do with my life. To my little brother, hanging out with you always makes me smile, and thanks for pretending I'm not that old.

Chapter 1

Introduction

As the ongoing global COVID-19 pandemic death toll hits 1.6 million, it has become clear that a better understanding of how diseases spread is critical to our global health and economy. The increasing availability of computational power and data on human mobility such as road and air traffic have driven advances in understanding and forecasting disease spread on a global or even metropolitan scale [1, 2]. However, we have not seen the same progress in understanding and forecasting disease spread on a local level. Our understanding of how diseases such as COVID spread through local institutions such as schools and offices, or even towns, has made minimal progress over the last ten years compared to our understanding of disease spread on a macro level.

One key obstacle is that while data on human mobility has increased our understanding of how humans interact on a global level, we have not had the equivalent data revolution on the local level. That is, we have not had an influx of data on our everyday face-to-face interactions. Measures such as social distancing, contact tracing, and restricted indoor capacity limits are all designed to understand and influence our face-to-face interaction times and distances, and have all proven to be critical in combating our current pandemic, and likely future pandemics.

Epidemiologists have long been interested in better understanding face-to-face interaction networks, times, and distances [3]. In addition, researchers and policy makers are often interested in contextual information that can greatly enhance the usefulness of interaction data, such as when people begin showing symptoms of being sick and people's health behaviors.

Typically, face-to-face interaction data is collected using interviews, specialized surveys, or other forms of self-reporting. While this type of data collection is flexible and allows epidemiologists and others to gather critical contextual information, the resulting data relies on the subjectivity of human memory, resulting in data of dubious quality. People may forget with whom they interacted, when they interacted with someone, or whether or not they interacted with particular people. In addition, it is unlikely people can recall exactly how long they interacted with someone, or exactly how far apart their interactions were. This is particularly true for fleeting interactions, such as hallway, water cooler, or bathroom interactions. Finally, manual contact tracing involves a significant amount of work for both the people being interviewed and the epidemiologists collecting the data.

In response to these issues, researchers have sought for decades to provide more objective, automated methods of gathering this data. One commonly explored method is wearable sensor systems, ranging from custom sensors to ordinary smartphones, to detect interaction times and distances between people carrying or wearing these sensors. Researchers and industry stake holders have also worked on sensors to objectively collect contextual health information. In particular, there have been efforts to detect when people cough, and in sensing hand-washing rates. Although there have been significant advances in these areas, by and large these systems provide either high-resolution data or practical usability, but not both.

1.1 Wearables for Face-to-Face Interaction Sensing

Wearables for face-to-face interaction can use a variety of technologies, including radios, ultrasonic transducers, and infrared sensors to detect when people are interacting. Ideally, these sensors would provide high-resolution interaction distances and times, while also being practical to deploy in a wide variety of situations.

The simplest and most common systems use RF proximity sensing. Using this technique, one wearable transmits a radio packet, which nearby wearables receive. In addition to getting the data in the radio packet, the receiving wearable's radio also reports a received signal strength indicator, or RSSI. Roughly speaking, RSSI increases as the distance between the transmitter and receiver decreases. Some systems simply use packet reception as a sign that two people are in close proximity, while other systems attempt to determine a "threshold" RSSI, with RSSI's above the threshold indicating that the two wearables are at most X m apart, where X is commonly set to 2 m. However, RSSI is affected by a myriad of other factors, and has not been shown to be a reliable metric across different times and spaces. Thus, in practice, it is difficult to determine exactly what range of interaction distances are captured, much less determine actual interaction distances. Mapping proximity data to interactions is further complicated by the fact that radio packets go through walls and tend to be omni-directional. The advantage of these systems is that they are easy to deploy. RF proximity sensing is relatively low power, and these sensors typically are able to operate without any dedicated infrastructure while maintaining a battery life of at least a day.

Other systems use more advanced techniques to estimate actual distances between wearables. Some still use purely radio based techniques, but with higher-power and more advanced techniques that allow them to more accurately estimate distances. Wearables that incorporate ultrasonic or infrared components can also ensure that the people wearing them actually have line of sight to each other, and are not just standing on opposite sides of a wall. While these systems can provide high-resolution interaction times and distances, they are significantly more power hungry, and typically require either dedicated infrastructure or extremely large batteries to operate. Thus, they are impractical to deploy in many cases.

1.2 Sensing Health Context

In addition to detecting interactions, contact tracing and other health related applications often require or are greatly enriched by a deeper understanding of peoples' symptoms and health behavior. Two important pieces of contextual information are understanding when people start coughing and how often people wash their hands, due to the ubiquity of coughs as a symptom in many illnesses and the importance of hand-washing to combating many types of diseases.

Detecting coughs is traditionally done by having people either fill out periodic surveys, or by having people self-report symptom onsets. Similarly, gathering data on hand-washing is typically done by surveys, self-reporting, or using hidden observers. These methods require significant manual effort, and can result in low-resolution or inaccurate data due to their reliance on memory and self-reporting compliance.

These issues have led researchers and industry stakeholders to explore automated cough detection sensors and hand-washing monitors. Cough detectors typically work by continuously recording a person's audio environment and detecting coughs in post processing. While this method has shown promising results, it results in a massive privacy burden on users, making it difficult to deploy in many contexts. Academic work on hand-washing sensors has focused on using accelerometers in smart-wrist bands to detect hand-washing motions. These systems have so far proven to be prone to false positives, and are generally untested in real-world scenarios. Commercial solutions have largely focused on using smart badges and soap-dispenser mounted infrastructure to detect hand-washing events. These systems are focused on increasing hand-washing compliance in hospital settings rather than accurate sensing. It is unclear how accurate these systems are, and their tailored use case can make them difficult to use outside of hospital settings. Furthermore, they are only able to detect hand-washing events involving soap, which paints an incomplete picture of hand-washing behavior outside of hospitals.

1.3 Active Sensing Limits Usability

The key problem that all of these sensors share is that actively sensing the desired phenomenon requires a high power budget, but wearable sensors are inherently limited in battery size. Getting high-resolution face-to-face interaction data involves utilizing high-powered radios to discover nearby wearables, synchronizing with them, and finally performing a ranging protocol to get interaction-distance. Detecting coughs in an audio stream requires significant computational power, which translates to a high power draw. Smart badge based hand-washing sensors require some form of localization to detect when users are near a sink, which again takes power to do accurately.

To compensate for the limited battery size a wearable sensor can support, many system designers offload work from the wearable sensor. Face-to-face interaction sensing systems may require infrastructure nodes to assist the wearables, and most cough sensors simply

record all of a user's audio, doing all the heavy computational lifting in post-hoc analysis. Otherwise, providing high-resolution data in these domains significantly limits battery life. Either way, these systems must compromise usability for high-resolution data. Without these compromises, these systems can only provide low-resolution data, if they can operate at all, or unacceptably low battery life.

We liken the current approach of active sensing to polling based architectures, where systems must constantly and actively poll for the data they want. Instead of further advancements in this architecture, we propose utilizing a pulling based architecture, where the high-powered components of the sensor stay asleep, only being woken up when a phenomenon of interest presents itself.

1.4 Thesis Statement

We claim that by utilizing low-power frequency and amplitude selective filtering in the acoustic, ultrasonic, and accelerometric domains, we can create hardware triggers (generically "passive vigilance") that enable the requisite power savings for "wearables" and "nearables" to achieve the size, weight, and power needed for practical multi-day deployments while gathering high resolution data on face-to-face interactions, coughs, and hand-washing behavior.

1.5 Contributions of this Dissertation

We have designed and implemented novel wearable sensors to collect high-resolution face-to-face interaction data and important contextual data without compromising usability. We do this by utilizing novel hardware-triggers and protocols in these spaces to construct passive-vigilance based sensors instead of the typical continuous active sensing architecture other sensors in these domains use. We believe these sensors not only advance the state-of-the-art in their respective applications, but also provide a path forward for future researchers to develop complex wearables without compromising usability. Although the systems we have built have a unified motivation of enabling a better understanding of disease-spread, they span three different applications. Therefore, we forego the traditional unified motivation and background chapter, instead splitting up the motivation and background into each individual project's chapter.

The foundation of this dissertation is Opo, a wearable sensor that can accurately track face-to-face interaction times and distances without compromising battery life or requiring dedicated infrastructure. Opo provides interaction distances up to 2 m and accurate to within 3 cm every 2 s while sporting a multi-day lifetime on a rechargeable battery the size of a Starburst candy. The initial conception of Opo was presented at Sensys'14 [4], although it has undergone significant hardware and protocol revisions since then. To the best of our knowledge, Opo still provides best-in-class face-to-face interaction time and distance resolution for its power budget.

Chapter 2 lays out the motivation and design of Opo. It revisits the history of face-to-face interaction sensors, identifying the fundamental architectural barriers that limited the battery life and usability of prior systems. We then present and evaluate the design of Opo, which utilizes a novel passive-vigilance based ranging protocol to enable high-resolution interaction sensing in a low-power, infrastructure free manner.

Chapter 3 explores the real-world benefits and limitations of using Opo to inform disease forecasting models. Because previous data sets that provide high-resolution interaction times and distances do not exist or are not available, it is unclear if this data can add significant value to forecasting or understanding disease-spread, or if Opo's high-resolution data can be inferred from low-resolution sensors. To explore this problem, we conduct a real-world deployment in a Midwest office setting in conjunction with epidemiologists at the University of Michigan, Ann Arbor and the University of North Carolina, Chapel Hill. Portions of this work are scheduled to be published in PLOS ONE [5].

In addition, we explore a subtle but fundamental problem for face-to-face interaction sensors. To the best of our knowledge, no face-to-face interaction sensor directly senses face-to-face interactions. RF based sensors use omni-directional barrier-ignoring distance or proximity as a proxy for face-to-face interactions, while chest-worn directional sensors such as Opo assume that people orient their bodies towards people they are interacting with. We explore how well chest-worn directional sensors such as Opo model face-to-face interactions by heavily instrumenting a shared-lab space with a motion capture system and conducting a short pilot study with Opos in this space.

Chapter 4 extends the Opo interaction-sensing sensor to also sense hand-washing rates. We overview current academic and industry solutions in this space, identifying a niche for a smart-badge + soap-dispenser sensor based system that can detect hand-washing instances, durations, and whether or not soap was used. To the best of our knowledge, our system is the only one that provides all three of these features. To validate our system's functionality, we conduct a short pilot test during the Midwest office deployment described above. Portions of this chapter are also scheduled to be published in PLOS ONE [5].

Finally, Chapter 5 utilizes the passive vigilance architecture of Opo to design a novel, privacy preserving cough sensor. We overview both prior cough sensors and related works on coughing as a symptom. While prior works focus on providing accurate cough counts, we find that there is significant value in simply knowing if people are generally coughing more. To fill this niche, we design a novel cough sensor that neither constantly records audio or actively analyzes an audio stream. Instead, we design a simple frequency and amplitude based hardware cough trigger that records a 1 s snippet of audio when it hears a sound that is potentially a cough, and is extremely unlikely to be triggered by vocal audio, thus preserving a user's privacy.

Chapter 2

Tracking Face-to-Face Interactions

2.1 Motivation and Background

Face-to-face interactions are an integral part of our everyday lives, playing a critical role in our mental and physical health, along with our professional productivity. Researchers have found that even without knowing what is said during these interactions, a better understating of our face-to-face interaction networks, times, and distances can help us better understand topics as varied as stress levels [6], racial attitudes [7], professional productivity [8], communication effectiveness and behavior [9], and differences in cultural norms [10, 11]. For this dissertation, we are motivated by the need to better understand how many diseases spread at a local level.

In particular, we are interested in a better understanding of how respiratory diseases such as influenza, SARS and COVID-19 spread. These diseases often spread through droplet or aerosol transmission, which are tiny drops of water that we naturally expel while breathing, coughing, and sneezing. The closer two people facing each other are, and the longer they interact for, the more of each other's droplets they are exposed to, and the higher chance they can infect each other if one of them is sick [12]. Numerous studies have shown that interaction time and distance both a play a role in infection rates for many respiratory diseases, and we have seen this topic come to the forefront of the COVID-19 pandemic with measures such as social distancing [13–16].

However, our understanding of everyday face-to-face interaction networks, times, and distances is limited. Face-to-face interaction data is traditionally collected using surveys, diaries, or contact-tracing interviews. These methods are flexible, and allow researchers to also collect contextual information. However, these methods also produce low-resolution data, and rely on the subjectivity and accuracy of human memory. It is unreasonable to expect people to remember precise interaction times or distances, and participants in these studies are typically only asked to recall people they have had contact with over a specified time period [17]. Even if we ignore interaction time and distance, people may simply not remember short-duration contacts [18]. These issues have driven researchers to develop objective and high-resolution interaction sensors [19–21]. The two most popular sensor solutions are Bluetooth proximity

sensing smartphones [22–27] and wearable RF proximity sensors [28–31].

Smartphone systems use Bluetooth Low Energy (BLE) scans to determine if two people are within close proximity of each other every 5 minutes. In these systems, smartphones transmit BLE packets to each other every 5 min, and if one smartphone can receive packets from another smartphone, then the two users are determined to be in close proximity. Wearable RF proximity sensors are technically very similar, but often scan every 20 s and use slightly different radios. In general, these systems define close proximity as within either 2 m or 5 m. To better determine if users are in close proximity, both types of systems may attempt to set a threshold received signal strength indicator, or RSSI. The RSSI of a received packet is a measurement of the power present in the received radio signal that contained the packet. In theory, RSSI has an inverse square power relationship with distance. Therefore, it should be possible to set a threshold such that RSSIs above the threshold mean users are within 2 m of each other, while RSSIs below the threshold mean users are farther apart. However, in practice, RSSI is affected by a multitude of dynamic environmental factors that can either attenuate or amplify RSSI values. Because of this, it is extremely difficult to use RSSI to determine distance or effectively set a threshold RSSI for real-world conditions [32]. Finally, due to their reliance on radios, these systems work through walls and do not take into account if people are actually facing each other, both of which are important considerations for disease spread [20, 33]. Optimistically, without the help of infrastructure nodes, RSSI proximity sensing is only accurate to within ± 2-3 m [34]. Nonetheless, the ubiquity of smartphones have led many countries to attempt to perform COVID-19 related contact tracing and exposure notifications using Apple and Google's contact exposure Bluetooth specification, which relies on BLE proximity sensing. These systems aim to detect when people are within 2 m of each other, and alert people who have been within 2 m of someone who contracts COVID-19. A recent evaluation of the Swiss, German, and Italian detection systems in a light-rail tram found that their performance "is similar to that of triggering notifications by randomly selecting from the participants in our experiments, regardless of proximity" [35]. Both this study and a similar study on a commuter bus [36] found that distance and BLE RSSI values are only weakly correlated. This matches our general intuition that RSSI values in these environments should be highly affected by multipath propagation due to buses and trains being made of metal, an RF reflective material. While most indoor environments are not encased in a metal shell, RF reflective surfaces and multipath propagation in regular indoor environments is extremely common [32, 37, 38].

The NIST TC4TL challenge aimed to utilize machine learning techniques to translate BLE RSSI values into distances [39]. This challenge provided sets of BLE RSSI values, along with a true distance measurement of either 1.2 m, 1.8 m, 3.0 m, or 4.5 m for each RSSI value set. The goal of the challenge was to correctly classify a BLE RSSI set to its true distance measurement while minimizing the combined false positive and false negative rate. Even the best scoring team had a combined false positive and false negative rate of 0.68 when attempting to correctly classify 1.2 m interactions, and on average had a combined false positive and false negative rate of 0.55 [40]. Overall, these results highlight just how difficult it is to use RSSI values to sense proximity, much less interaction distance.

System	Ranging Method	Ranging Accuracy	Infrastructure	Time Resolution	Size	Battery Size	Battery Life	Tested on People
Opo	**UL/RF TDoA**	**5 cm**	**No**	**2 s**	**14 cm^2**	**40 mAh**	**93 hr**	**Yes**
Socitrack [a][41]	UWB ToF	15 cm	No	2 s	21 cm^2	2000 mAH	120 hr	Yes
WREN [28]	RF Scan	200 cm	No	20 s	13 cm^2	180 mAh	16 hr	Yes
TelosB [31, 42]	RSSI Sensing	200 cm	No	20 s	20 cm^2	4000 mAh [b]	16 hr	Yes
Social fMRI [43]	Bluetooth Scan	500 cm	No	300 s	N/A	N/A	N/A	Yes
WASP [37]	RF ToF	50 cm	Yes	.04 s	N/A	6.5 Ah	10 hr	Yes[c]
Cricket [44]	UL/RF TDoA	10 cm	Yes	1 s	40 cm^2	4000 mAh [b]	N/A	Yes
iBadge [45]	UL/RF TDoA	10 cm	Yes	N/A	38.5 cm^2	N/A	5 hr	Yes
Dolphin [46]	UL ToF	24 cm [e]	Yes	13 s	N/A	N/A	N/A	No
Future UL [47]	UL AoA, ToA	sub-cm	Yes	1 s	N/A	N/A	N/A	No

[a] Socitrack has configurable time and distance resolution which determines its battery life.
[b] Systems used 2x AA batteries. Listed mAh is average mAh of two alkaline AA batteries.
[c] WASP evaluated humans in outdoor settings, but not indoor environments.
[d] RADAR assumes access to a pre-existing wireless network, ability to build and access an RF map of a building.
[e] DOLPHIN accuracy under semi-ideal, static conditions. Under ideal conditions, accuracy is 2 cm

Table 2.1: Comparison of Opo to systems representing various ranging techniques. Opo was the first infrastructure-free system that could character interactions at a sub-meter level. To the best of our knowledge, it is still the only one that can do so while taking into account if people are facing each other and ensuring that people are not separated by walls.

Motivated by the need for a sensor that can provide high-resolution interaction times and distances, in this chapter we survey prior work on interaction sensor technologies and present and evaluate the design of Opo, a high-resolution, infrastructure-free, wearable interaction sensor. Opo provides an average interaction distance error of 5 cm, an interaction time resolution of 2 s, and a 90 hr battery life on a 40 mAH hr battery the size of a Starburst candy.

2.2 Related Work

While real world interaction sensor deployments have universally used some form of RF or BLE proximity sensing, a variety of ranging and localization techniques can be used to perform contact-tracing. Here we survey these techniques, finding that our surveyed systems can provide either high-resolution interaction data or usability, but not both, and explore why that is. Table 2.1 provides an overview of Opo and other systems surveyed here.

Pure RF Techniques and Systems

Pure RF ranging techniques are appealing since an RF channel is often needed as a data link regardless of how sensors detect interaction distance and time. The simplest technique, Bluetooth/RF proximity scanning, has already been discussed above. In these systems, sensors or smartphones periodically send out wireless packets in the 2.4 GHz ISM band, and the reception of the packet or the RSSI of the packet is used to determine if users are in close proximity. Many of these systems claim that they can determine if users are within 2 m of each other, but these claims are based on limited and controlled testing scenarios [28, 31, 42, 48]. More comprehensive testing has shown that this technique is highly unlikely to be

able to make this determination during real-world deployments [32, 35]. More realistically, these systems can only determine if people are within 4-5 m of each other [32]. We note that although the OpenBeacon/Sociopatterns sensor is often titled as an RFID sensor, it is actually a BLE beacon [48]. RADAR and other

An improvement over RSSI measurement is coordinated pairwise RF time-of-flight (RF ToF) ranging, a peer-to-peer approach supported by systems such as Waldo [49] and WASP [37]. RF ToF works by measuring the round trip time (RTT) of a radio packet between two nodes. As an oversimplification of how these systems work, one radio sends a packet to another radio at a very precise time, and the receiving radio records precisely when the packet arrives. The receiving sensor then calculates the time-of-flight (ToF) of the packet by subtracting the receive time by the transmit time, and multiplies the ToF by the speed of light to get its distance to the transmitter. Measurement accuracy is heavily affected by multipath propagation, which is prevalent in indoor environments [37]. Using traditional narrow-band radios, such systems are accurate to within 0.5 m 65% of the time in static conditions not involving humans [37].

Traditional narrow-band RF ToF systems are power hungry due to the high-speed processing needed to range with sub-meter resolution, requiring up to 2.5 W for transmissions and 2 W for receptions [37]. In addition, these systems typically require pairwise ranging, which scales poorly with study size, requiring $O(n^2)$ ranging operations for a group of n individuals. Furthermore, these systems must schedule collision free ranging operations in a dynamically changing network, which is a significant challenge in power constrained environments such as wearable sensor systems. Pairwise ToF ranging is also possible using acoustic/ultrasonic channels, but these systems suffer from the same scaling and scheduling problems as RF ToF systems.

However, recent advances in ultra-wideband (UWB) radios, neighbor discovery protocols, and wireless synchronization protocols, have enabled the creation of much more usable and accurate RF ToF interaction sensors. Ultra-wideband radios can compensate for multipath propagation by utilizing multiple antennas and frequencies, resulting in much more accurate real-world RF ToF ranging accuracies compared to traditional narrow-band RF ToF [50]. Biri et al utilize advances in all these fields to develop SociTrack, an infrastructure free UWB ToF wearable interaction sensor [41]. Socitrack is accurate to within 15 cm, and can provide a 2 s time resolution. Although Socitrack still draws about 40 x the power of Opo, it can also be kept in a pocket while Opo must be attached to a lanyard or pinned to a shirt. This means that in practice, Socitrack can often support a much larger battery than Opo can. However, like other RF based systems, Socitrack ignores walls and other barriers and whether or not users are facing each other. For their use cases, these are positive features. For the purpose of understanding disease spread though, we want to know how long people are facing each other for, and only when they are not separated by barriers.

TDoA Systems

Time-difference-of-arrival (TDoA) systems require both speed of light (RF) and speed of sound (ultrasonic) hardware frontends, but provide higher distance accuracy than purely radio based techniques, commonly achieving accuracies under 10 cm. In addition, the ultrasonic component of these systems is directly and does not go through barriers, meaning these systems only work when users are facing each other and not separated by a wall. TDoA ranging generally works by sending out a radio packet and an ultrasonic or audio pulse at the same time. The radio packet travels at the speed of light, effectively arriving at a reciever instantaneously. In contrast, the ultrasonic/audio pulse travels at the speed of sound, and arrives at the receiver at a later time. By taking the time difference of arrival between the radio packet and ultrasonic/audio pulse and multiplying it by the speed of sound, a receiver can trivially and accurately calculate its distance to a transmitter.

Smartphones can perform TDoA ranging in the audio domain without infrastructure nodes, and are accurate to within 5 cm [51]. However, this requires that the phone be held in certain orientations due to audio directionality and does not work while the phone is in a pocket or bag. Thus, smartphone TDoA is much more suited applications such as indoor GPS rather than interaction sensing [52, 53].

Wearable sensor systems such as Dolphin and Cricket implement TDoA ranging using radios and ultrasonic (UL) hardware to avoid user annoyance [44, 54]. These wearable sensor systems are indoor localization systems, and utilize infrastructure nodes both to localize the wearable sensors and to remove a lot of design complexity and power burden from the wearables. Infrastructure nodes can effectively relax the $O(n^2)$ message complexity of pairwise ranging to $O(n)$ by serving as constant beacons to mobile receivers. Receivers can then use the transmitted information to estimate their location and orientation. Custom fabricated broadband ultrasonic systems have been able to provide sub-centimeter localization under heavily controlled environments, with bulky and unwearable nodes [47]. Among infrastructure systems that can be deployed today, it is reasonable to expect up to 10 cm spatial accuracy [44]. Some systems, such as DOLPHIN, report accuracy as high as 2 cm, but only under ideal, static conditions [46].

A different approach to addressing the $O(n^2)$ message complexity of pairwise ranging, without the overhead and co-location errors of infrastructure-based approaches, is to use broadcast (one-to-all) ranging with TDoA between RF and ultrasonic pulses. AHLoS [55], and its successor iBadge [45], offer a decentralized, ad-hoc broadcast ranging scheme. But, AHLoS relies on the underlying synchronization provided by a DSDV [56] variant that is ill-suited to dynamic mobile networks. Hence, it is reduced to a static-plus-mobile system with infrastructure-dependent synchronization of ranging events. Cricket [44] also employs broadcast ranging. However, it too utilizes infrastructure nodes to remove the burden of neighbor discovery and scheduling from the wearable sensors.

Part of the problem with these systems is exactly how much infrastructure is required. Even deployments which require just a single visible infrastructure node have proven problematic [57]. Many of the above systems are localization systems where the wearables only range with

infrastructure nodes. While this reduces the power burden on the wearable, this requires at least three visible beacons to localize. This is especially problematic for ultrasonic beacons, since they require line of sight to minimize error. The Active Bat system, a RF/UL TDoA localization system, found that 100 infrastructure nodes are required to achieve sufficient coverage in a 280 m^3 office space. Assuming a 3 m ceiling height, this means that they require over 1 node per square meter of ceiling [58].

For a general-purpose interaction sensing system, infrastructure imposes a significant usability barrier. Installing and maintaining infrastructure is laborious and expensive, and limits researchers to characterizing interactions in specific settings rather than characterizing interactions between specific people. Even when resources are available to instrument a building and only one infrastructure node needs to visible to the wearables' radios, ensuring coverage can be difficult. Isella et al conducted a study of 195 individuals rotating through a single hospital ward using an RF interaction sensor with such requirements. The authors found that 69 of the individuals were missing over 25% of the expected data, and blamed problems with ensuring strong radio coverage for the missing data [57]. While infrastructure does not make a system impossible to deploy, we note that the largest interaction sensing deployments we can find all use infrastructure-free systems [22, 28, 31].

Discovery, Synchronization, and Wakeup

Our survey of systems found that in general, localization and interaction sensing systems trade off data resolution and usability. Most systems either push much of the power burden to infrastructure nodes and are able to provide high-resolution interaction times and distances, or are infrastructure free, but only provide limited interaction time and distance resolution. The key problem is that in addition to the power required to actually perform a ranging operation, wearables must also support a neighbor discovery protocol to discover other nearby wearables, and a scheduling protocol to schedule ranging operations with these nearby wearables. Since the creation of many of these systems, various techniques have been proposed to discover and synchronize mobile nodes with duty-cycled radios, including periodic extended transmission [59], listening [60], random [61], and deterministic [62, 63] neighbor discovery.

However, these advances are not a panacea for past systems. These protocols expose a fundamental power vs latency tradeoff that requires interaction sensing deployments to select between battery life/form factor and temporal fidelity. In the end, with traditional neighbor discovery and scheduling protocols, the faster we want to discover nearby nodes and schedule ranging operations with them, the more often we have to turn a sensor's processor and radio on. Even with advances in microprocessor and radio technologies, these are still high-powered components for a wearable sensor. For example, in preliminary, controlled tests, UConnect, a low-power neighbor discovery protocol, offers a 2.5 s discovery latency with a 1.3 mW power draw, which is enabled by using 250 *mus* discovery slots, instead of the 1-2 ms slots common among other protocols [62]. In contrast, a 7-day real-world pilot deployment with Opo showed a 1.41 mW power draw using the same radio while providing TDoA ranging estimates with fewer RF messages.

Wakeup radios sidestep the power/latency tradeoff inherent in discovery and synchronization protocols. A survey of wakeup radios reveals 20 designs that draw 50+ μW, with receive sensitivity roughly between -100 and -20 dBm, and one design that draws 98 nW with a sensitivity of -41 dBm [64]. However, wakeup radios are still a nascent technology requiring custom fabricated chips, and it is unclear when they will be ready for mass use. Ultra low power custom fabricated ultrasonic wakeup radios also exist, but again are not available commercially and are not validated outside of lab settings [65, 66]. Opo builds on this work though by multiplexing a single transducer for ultrasonic transmissions and receptions, demonstrating viability with commercial components, improving noise sensitivity, and using ultrasonic wakeups to trigger scalable TDoA ranging.

2.3 Design Requirements and Targets

It is unclear exactly what level of interaction distance resolution is required to better understand disease spread, but work on aerosol and droplet expulsion while breathing and coughing suggests that at a minimum sub-meter ranging accuracy would be beneficial, and that higher accuracies may still provide value [67]. At the very least, the higher our distance resolution, the better we can determine if participants are within 2 m of each other, which is the goal of many deployed interaction sensing systems [28, 31, 35]. 2 m is also the recommended interaction distance to significantly reduce the chance of infection from many respiratory diseases such as COVID-19 [16]. Because of this, we also set the maximum interaction distance we are interested in to 2 m, which is in line with previously deployed interaction sensing wearables [28, 31]. Previously deployed interaction sensing wearables commonly offer a 20 s time resolution, but greater temporal resolution may be significant [68]. We thus set a minimum bar of 20 s temporal fidelity.

Because high-resolution interaction distance and time data does not exist, it is difficult to know exactly what level of interaction distance and time resolution matter. Overall, our goal is to maximize our interaction time and distance resolution without breaking the usability and scalability requirements discussed below.

Defining Usability

Usability, or the ease and comfort at which a system can be deployed by researchers and used by participants, has proven to be important to an interaction sensing system's real-world prevalence. To set usability requirements, we draw upon past wearable sensor deployments.

Researchers from the University of Utah successfully deployed wearable pure RF proximity sensors called WRENs to over 8,000 school-age children [28]. The authors reported no issues of discomfort or other burdens from participants. We treat the volume and weight of the WREN (13 cm^2, 10.8 g) as an upper bound for our form factor. The WREN deployment [28] and past TelosB deployments [31] indicate that at minimum, a day-long battery life is needed,

although greater battery life is highly desirable. The WREN deployment specifically desired a week-long battery life [28].

The other key property we derive from WREN is the criticality of infrastructure-free operation. Without characterizing the physical space, it is impossible to know exactly how much infrastructure is necessary or where to best place it. Beyond the resource cost of set-up and tear-down, infrastructure-dependency also defines physical boundaries on interaction sensing, making systems less general. Anecdotally, we found in discussions with public health professors that even obtaining permission to set up infrastructure can be a formidable barrier.

For the design of a flexible and practical interaction sensing system, we consider infrastructure an untenable requirement to impose on Opo.

Scalability vs The Common Case

Since a circle 2 m in diameter can only fit 16 average U.S. adults around its perimeter [69], it is not necessary for 1,000 Opos to simultaneously communicate with one another. We optimize Opo for the common case: 2-10 individuals within 2 m of each other, and speculate on ways to scale Opo for hyper-dense scenarios in Section 2.7.

Interactions vs Proximity

Current interaction sensing systems often equate proximity with interactions. If a phone's Bluetooth scan spots another phone, we assume that two people are interacting, even if they are just working in neighboring offices. More advanced RF ranging also does not solve this problem, since people may in fact be in close proximity, just with a wall between them. Intuitively, this assumption is prone to false positives in many scenarios, such as office workers sitting on either side of a common cubicle wall. Furthermore, case studies in disease spread have shown the need for more directional interaction sensing in addition to better interaction distance sensing [20, 33]. Thus, we desire a system which captures interaction distances only when users are facing each other with no walls between them. UL/RF TDoA systems mitigate this problem by exploiting the inherent directionality of UL transducers and the inability of ultrasonic signals to propagate through physical objects such as walls. This makes UL/RF TDoA an attractive ranging prospect for us, and is one the reasons we built Opo around this ranging technique.

2.4 Design

Opo is designed to provide infrastructure-free interaction distance sensing with better temporal fidelity than past infrastructure-free TDoA systems. At a high-level, Opo achieves this by replacing active RF neighbor discovery and scheduling with passive vigilance in the ultrasonic domain. This is made possible by a novel ultrasonic wakeup radio, which enables an Opo sensor to use broadcast ranging to asynchronously wake up and range with all n neighbors in

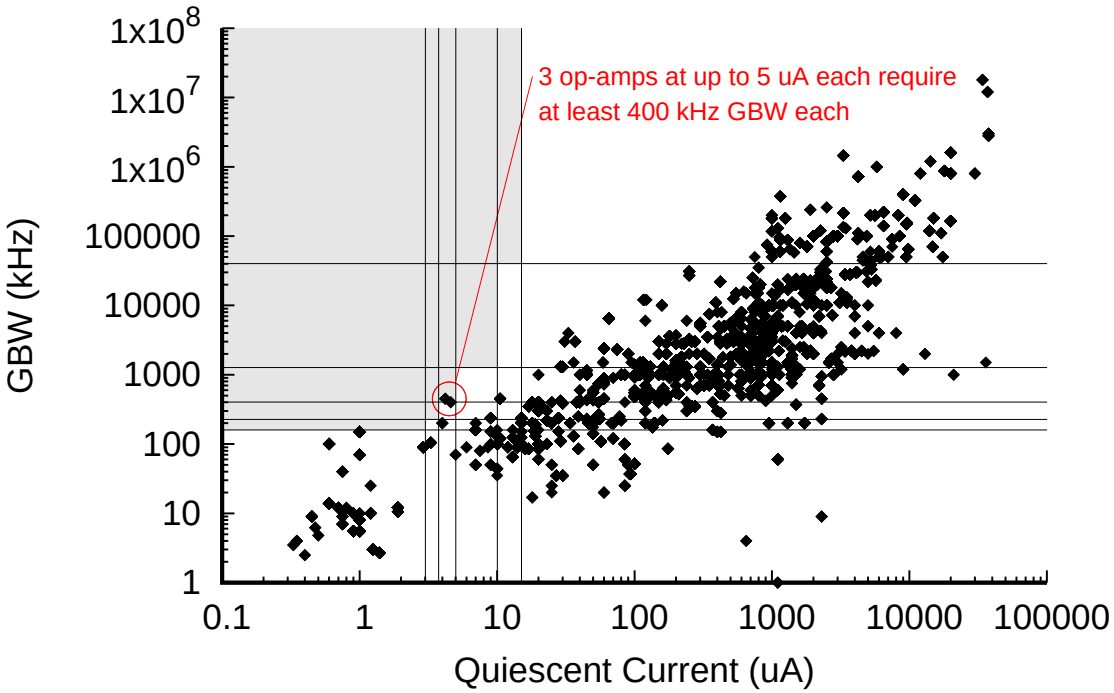


Figure 2.1: A survey of op-amps comparing power draw and gain. Cascading op-amps results in exponential growth in gain with only a linear growth in power. The shaded gray region contains the gain to power tradeoff from one to five op amps (left-top to right-bottom) normalized against the circled op amp we chose.

one interaction. We first introduce our broadcast ranging primitive and then build up our Opo design, starting from the ultrasonic wakeup frontend building block into a complete system.

Asynchronous, Broadcast Ranging

Coordinating wakeups and communication in a highly mobile network is extremely challenging. Traditional neighbor discovery protocols are not designed to be run every second and a centralized architecture violates our infrastructure requirement. Opo solves this problem by avoiding coordination and bi-directional communication.

Opo sensors have an always-on low-power ultrasonic wakeup frontend. An Opo ranging operation begins when one sensor elects to announce its presence by transmitting an ultrasonic pulse. Any sensor that receives this wakeup pulse prepares to receive a range event from the transmitter. A range event is a unidirectional transmission from the transmitter to all n neighbors that receive the signals. At the end of a range event, all receivers know their range from the transmitter. By permitting this asymmetry and avoiding bidirectional communication, Opo enables 1::n ranges to be captured per range event, enabling a network

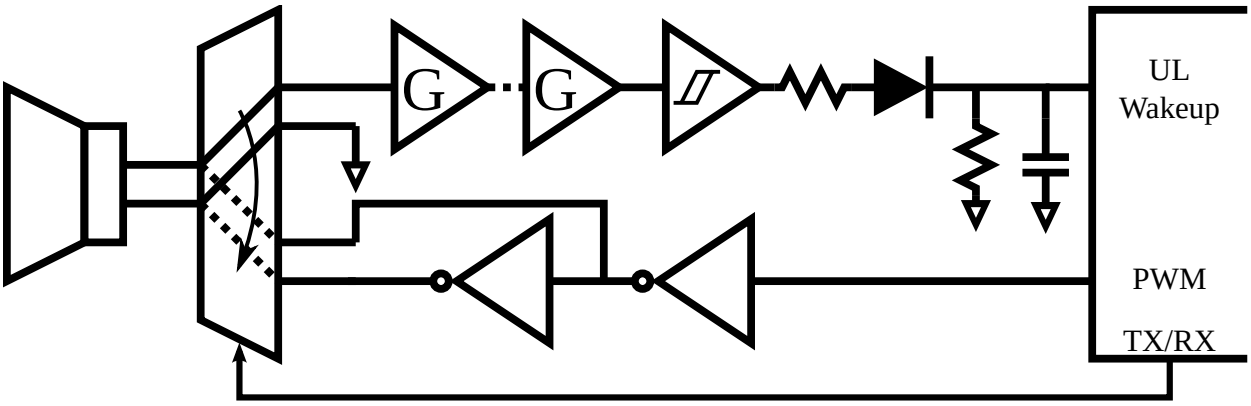


Figure 2.2: The Opo ultrasonic subsystem is composed of a low-power receive frontend (top) and a simple inverter-driven transmit frontend (bottom). An analog switch controls whether the transducer is transmitting or receiving.

of n sensors to calculate all of their mutual ranges in only n range events as opposed to the n^2 events required in pairwise ranging schemes.

Exponential Gain = Low-Power Wakeup

The key to our broadcast ranging primitive is the always-on ultrasonic wakeup circuit. A piezoelectric ultrasonic transducer will generate a small amount of current when struck by an ultrasonic pulse (the Opo wakeup). The question is whether this tiny current can be reliably and accurately detected within a constrained power budget.

Empirically, we find that a 1000× gain[1] is required to reliably detect the ultrasonic signal. To do this, we scraped the product summaries Digikey provides for its catalog of op-amps. We surveyed 15,000 op-amps that can be run from a typical 3.3V power supply to determine how much power would be required to provide a 1000× gain. Figure 2.1 explores the tradeoff space between the number of gain stages, the gain of each stage, and the total current draw.

In theory, we can use any number of op-amps in our ultrasonic front end. In practice, we are highly motivated to limit the number of op-amps used. Since filters are imperfect, the noise floor increases exponentially with the number of op-amps, reducing the voltage area (VCC - noise floor) that our system has to work with. In early experiments, even using 4 op-amps required significant fine tuning and additional filtering in the system. Furthermore, adding additional op amps increases our form factor, limiting the wearability of our sensors. Based on these factors, we limit our search space to 5 op amps. Our survey found that we can achieve a $1000x$ gain with a 3 stage amplification front end that draws 19 μA when idle. Even using a small, 40 mAh battery, the idle listening current of our UL wakeup front end consumes less than .1% of our power budget over a week long lifetime.

[1]This is same gain required in SpiderBat [70].

Ultrasonic Frontend

Figure 2.2 shows the Opo ultrasonic subsystem. The amplified ultrasonic signal is fed into a comparator and integrator which is connected to a MCU interrupt pin. The MCU also controls an analog switch that can connect the ultrasonic transducer to a transmit frontend for pulse generation.

Ranging: RF and Ultrasonic TDoA

Opo uses UL/RF TDoA ranging, which as discussed in Section 2.2 is accurate, directional, and does not penetrate walls, making it a natural fit for our design requirements. Opo's ranging primitive is uses the well-studied "thunder and lightning" approach. During a ranging event, the transmitting sensor simultaneously transmits an RF packet (speed of light) and an ultrasonic pulse (speed of sound). The transmitter includes a unique ID in the RF packet to identify who is transmitting the range event. In the same way that humans can estimate their distance to a storm by counting the time between a lightning strike and thunder clap, receivers can calculate their distance to the transmitter by counting the time between the RF packet arrival and the ultrasonic pulse arrival, provided that the ultrasonic and RF signals are sent simultaneously.

In practice, synchronizing the RF packet and ultrasonic pulse in software can be nontrivial, and each millisecond of desynchronization can result in a 0.3 m, or 1 ft error in distance estimates. Different software and hardware stacks may result in various delays between an application issuing a radio transmit command and the radio packet actually being sent, especially if the radio is kept off until transmission. Similarly, different hardware and software stacks may result in delays between transmitting an ultrasonic signal at the application layer and the resulting transmission at the hardware level. While synchronized transmission can be accomplished with a significant amount of low-level control, writing this code is often tedious and difficult, and results in a non-portable application.

To compensate for this, we design a ranging protocol that hardware-synchronizes the radio and ultrasonic transmissions, allowing us to implement ultrasonic RF TDoA ranging with minimal low-level code, as shown in in Figure 2.3. We assume that the radio exposes a start of frame delimiter, or SFD, interrupt or status bit that marks the actual start of the radio packet transmission. In this protocol, the transmitting Opo first sends out a UL wake up pulse. Upon receiving this pulse, nearby Opos wake up and turn on their radios. A short while later, the transmitter then simultaneously transmits a ultrasonic pulse and RF packet, which contains a unique id. Receivers record the time they receive the second UL pulse and the time they receive the radio packet, and multiply the difference in time by the speed of sound to calculate their distance to the transmitter. The transmitter hardware-synchronizes the ultrasonic and radio transmissions by adding an AND gate between the PWM signal and inverters in the ultrasonic transmission front-end shown in Figure 2.2. The AND gate is driven by the PWM signal and a GPIO line tied to the SFD interrupt which is idle low. The transmitter first begins generating the PWM signal before issuing a radio transmit command.

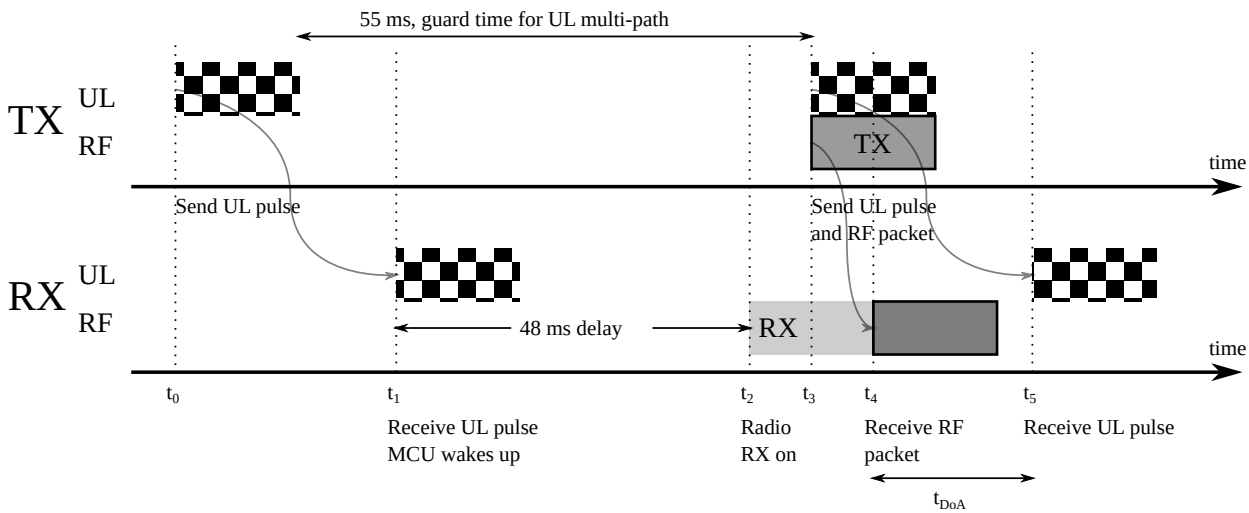


Figure 2.3: An Opo ranging operation begins with a wakeup UL pulse. Upon receipt, a receiver disables its UL RX frontend and starts a timer and waits for 48 ms for any UL multipath signals to dissipate. The receiver then enables its radio in anticipation of the TDoA signals and starts a 8 ms range-failure timeout. 55 ms after the wakeup pulse, the transmitter sends a simultaneous RF packet and ranging UL pulse. Upon receipt of the RF packet, the receiver re-enables its UL RX frontend and receives the ranging UL pulse. The receiver computes the TDoA between the RF and ultrasonic pulses to calculate its range from the transmitter. If the receiver's 8 ms range-failure timer triggers before receiving the complete range sequence, it abandons the range operation and returns to idle.

The SFD line is pulled high when the radio packet transmission actually begins, causing the PWM signal to propagate through the AND gate, synchronizing the radio packet and ultrasonic pulse transmission.

We also design a ranging protocol that uses only one ultrasonic pulse, but requires much more low-level control over the radio, as shown in Figure 2.4. In this protocol, the transmitter sends out one ultrasonic pulse, and then a short while later sends out a radio packet. As long as a receiver knows the precise time between the radio packet and ultrasonic pulse transmission, it can trivially calculate its distance to the transmitter. This requires a significant amount of low level code to either precisely transmit the radio packet a fixed time after the ultrasonic pulse, or to write the time difference of transmission to the radio packet as it is being sent. While this protocol does not significantly reduce the energy required to perform a ranging operation, it halves the amount of ultrasonic noise generated and reduces the total time it takes to perform a ranging operation by roughly 80%. This is because in the hardware-synchronized protocol the system must wait for ultrasonic multipath signals to die down, while in this protocol no such wait is needed.

Regardless of which protocol is used, the key idea here is that Opo sensors are normally asleep and passively listening for nearby Opos using our ultrasonic wakeup radio. In contrast,

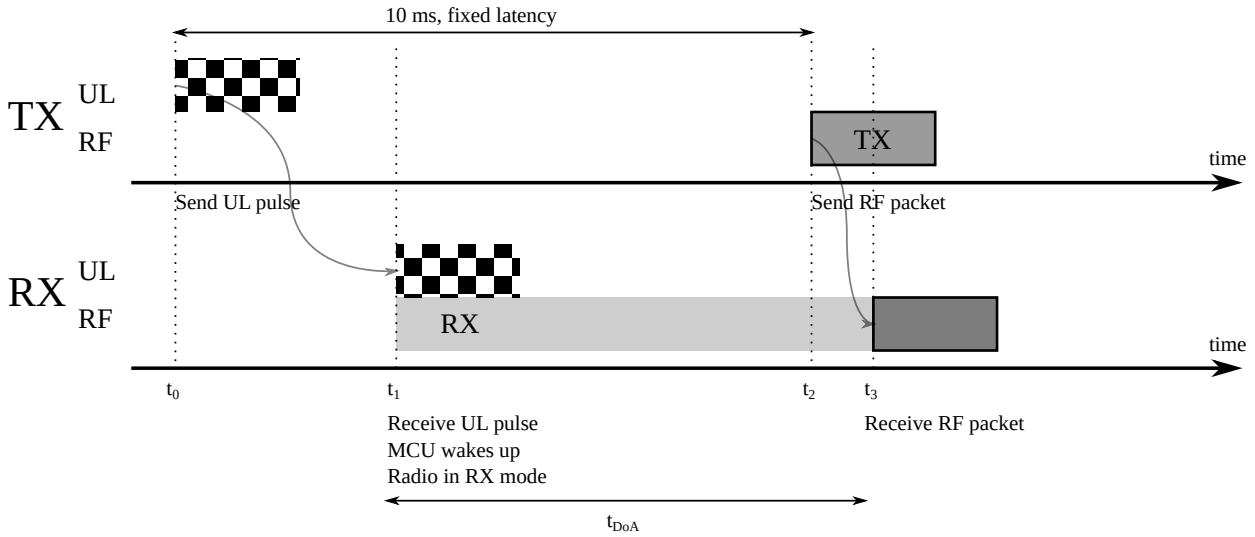


Figure 2.4: An optimized version of Opo's ranging protocol that requires more low-level radio control. The second ultrasonic pulse is eliminated, calculating the TDoA instead between the wakeup pulse and the subsequent RF packet. The delay for sending the RF ranging packet is also reduced (40 ms to 10 ms). A 10 ms ultrasonic wakeup window covers approximately 3 m of ranging. Sensors further away would miss the radio packet as they would wake up too late to receive it.

prior infrastructure free systems have to run an RF neighbor discovery protocol, which requires the sensor to periodically wake up both its micro-processor and radio even if no other sensors are nearby. This dramatically reduces Opo's power usage and simplifies Opo's software complexity.

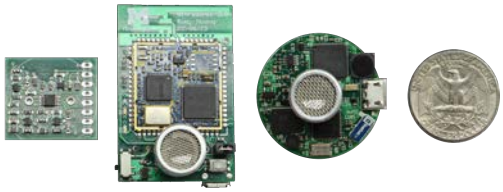
The Unimportance of Congestion Control

An obvious question is how an Opo sensor decides when to transmit and how it avoids TX collisions. In our protocol, each sensor simply waits for a random interval uniformly drawn from 1-3 s with ms granularity between each ranging operation. To provide some intuition for why this simple transmit scheme is sufficient, we examine the possible collisions. Since an Opo range operation takes only around 50 ms, we simplify our analysis by ignoring the effects of sensor mobility during a single ranging operation.

Directionality differentiates the broadcast domain of ultrasonic and RF signals. If two sensors are proximal but not facing each other and choose similar transmit times, a receiver may detect the wrong transmitter's RF packet, artificially increasing the TDoA from the wrong transmitter. For this collision to occur, the extra packet must arrive after the receiver's multipath delay and before the actual transmitter's RF packet—an 8 ms window at most. Collisions can also occur if a new transmitter elects to send its wakeup pulse between the RF and ultrasonic pulses of an ongoing ranging operation. This type of collision would cause

a receiver to record an artificially short range with the correct transmitter. A final type of collision may occur if two sensors elect to transmit at the same time. The impact of such a collision would be a missing ranging event for each sensor, as Opo sensors cannot transmit and receive at the same time.

We hypothesize that for reasonably sized groups, the probability of these collisions is sufficiently low and Opo's temporal fidelity is sufficiently high to compensate for the few errors that will occur.



(a) Various Opo implementations.

(b) Various different wearing scenarios of Opo.

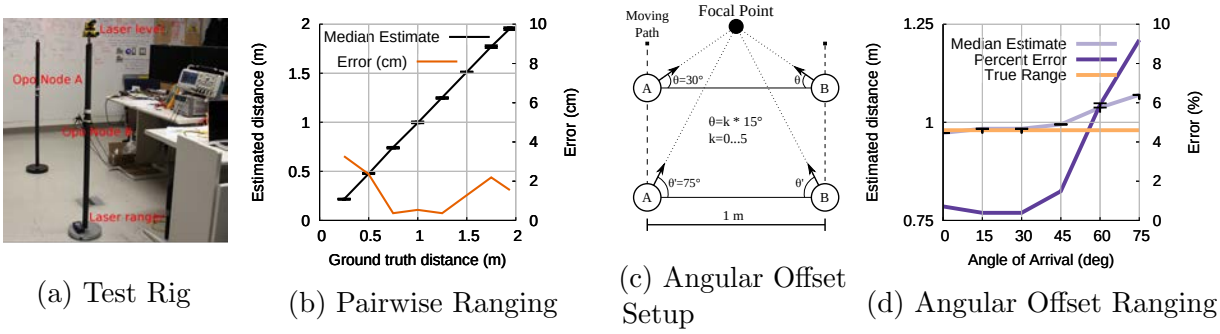
(c) Current Generation of Opo

Figure 2.5: Implementations and wearing scenarios of Opo. Figure 2.5a shows various implementations of Opo. The leftmost board is a standalone Opo frontend, designed to act as a daughter-board for standard mote platforms. The center board is the evaluated Opo system. The round board is optimized for form-factor, and can be worn in a case or by sticking the transducer though a button hole. Figure 2.5b shows various ways to wear Opo. The Opo in the upper left acts as a shirt button; the Opo in the upper right is a lapel pin; the Opo in the bottom left attaches to a tie clip; the Opo in bottom right is a badge clip. Finally, Figure 2.5c shows the latest iteration of Opo, which measures 1.3 x 1 in. This version has a BLE radio to enable smartphone communication

2.5 Implementation

To evaluate our design, we implement several generations of the Opo platform, refining the power, performance, and form factor(Figure 2.5a). We show our latest design of Opo in Figure 2.5c, which is what we use in Chapter 3 and Chapter 4. In Figure 2.5b, we show various ways that Opo can be worn.

We initially built our Opo prototype around the EPIC platform and TinyOS [71], which used to evaluate Opo in this chapter. Since then, we have transitioned to using a TI CC2538 and the Contiki operating system. From the three candidate op-amps in our survey (Figure 2.1), Micrel's MIC861/863 provides the best amplification to power tradeoff. We select the Prowave 400PT120 transducer for its tight 40 kHz center frequency. We empirically derive the transducer response time to establish a baseline offset and subtract 15 cm from raw TDoA range calculations. Our evaluation in Section 2.6 finds our constant offset to be consistent between transducers. Our schematic and software stack are available at github.com/lab11/opo. Our TDoA calculations assume the speed of light is instantaneous and the speed of sound to be 348.485 m/s.

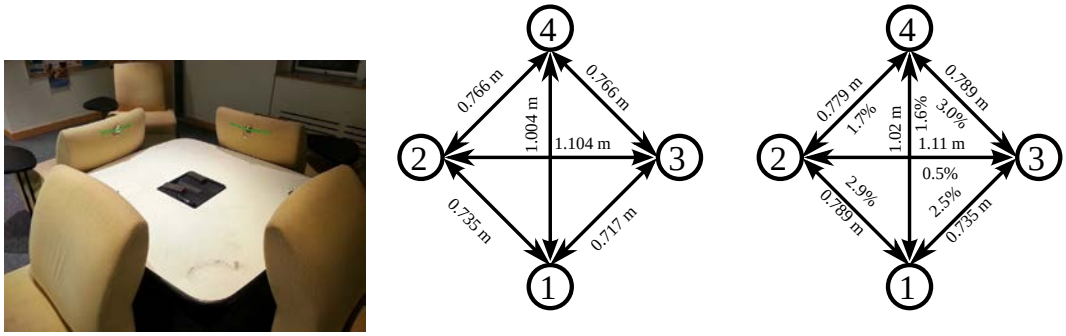


(a) Test Rig (b) Pairwise Ranging (c) Angular Offset Setup (d) Angular Offset Ranging

Figure 2.6: Controlled Microbenchmarks: Our test set up (2.6a) is comprised of Opo sensors mounted on two metal poles, a laser range finder, and a laser angle finder. 2.6b shows that our average error is usually < 2 cm. The black bars are the 95% error bars for Opo's range estimations, showing that Opo is both accurate and precise. To test angular offset performance (people chatting in a circle), we pointed two Opo sensors at the same focal point and slowly increased the angle between them, as shown in 2.6c. Our results (2.6d) show that while error does increase with angular offset, average error is only 6% at 60°(six people in a circle).

2.6 Evaluation

In our evaluation of Opo, we use the ranging protocol shown in Figure 2.3. We evaluate Opo's ranging accuracy, temporal fidelity, and power usage, along with the effects of density on Opo's performance with a series of controlled tests. We then evaluate Opo's ability to characterize



(a) Setup (b) Ground Truth Ranges (c) Opo Measurements and Percent Error

Figure 2.7: Four-Way Ranging Accuracy: Setup (2.7a), ground truth (2.7b), and Opo measurements (2.7c). The measurements shown in 2.7c are the median measurements between each pair of sensors, aggregated over the duration of the experiment (12 min).

human interactions with a series of human experiments, testing Opo's ability to capture both short and long interactions and individual and group interactions. We further show that Opo works for both sitting and standing interactions, and casual and work interactions. Following that, we inspect the energy usage of Opo and develop a power model to inform deployments. We finish our evaluation with a week long deployment with 8 participants to evaluate real-world performance and battery life.

Recalling that an Opo sensor records no information when it transmits, we evaluate Opo with a global perspective. We aggregate all of the collected ranges and consider pairwise metrics using this aggregate view, independent of which sensor was the transmitter or receiver.

Ranging Microbenchmarks

We use a test rig (Figure 2.6a) to conduct a controlled evaluation of ranging accuracy and the effects of angular offsets.

Pairwise Ranging Accuracy

To evaluate face-to-face accuracy, we perform Opo ranging measurements at a series of distances between 0.25 and 2 m, as shown in Figure 2.6b (with 95% error bars). We conduct 45 range measurements at each distance and find an average error of 0.015 m. For a given distance, 95% of ranging measurements are within 0.02 m of each other.

To evaluate precision across UL transducers, we take 10 transducers and conduct this test at 2 m, taking 45 measurements with each transducer. We find that 95% of all 450 measurements fall within 0.01 m of each other, with a median range measurement of 2.00 m.

Ranging Errors due to Angular Offsets

In group interactions, people often face the center of the group, leading to symmetrical angular offsets between people. We evaluate the effects of angular offsets on ranging accuracy by sweeping symmetrical offsets from 0° to 75° while keeping distance constant (Figure 2.6c). We conduct 10 ranging measurements at each angle, and find in Figure 2.6d that increasing offsets leads to an increase in measured range. Precision is unaffected, with 95% of ranging measurements at a given angle falling within 0.017 m of each other. Even at a 75° offset, Opo is accurate to within 0.01 m, a generational improvement over the 2 m resolution of RF scans.

Four-Way Ranging Accuracy

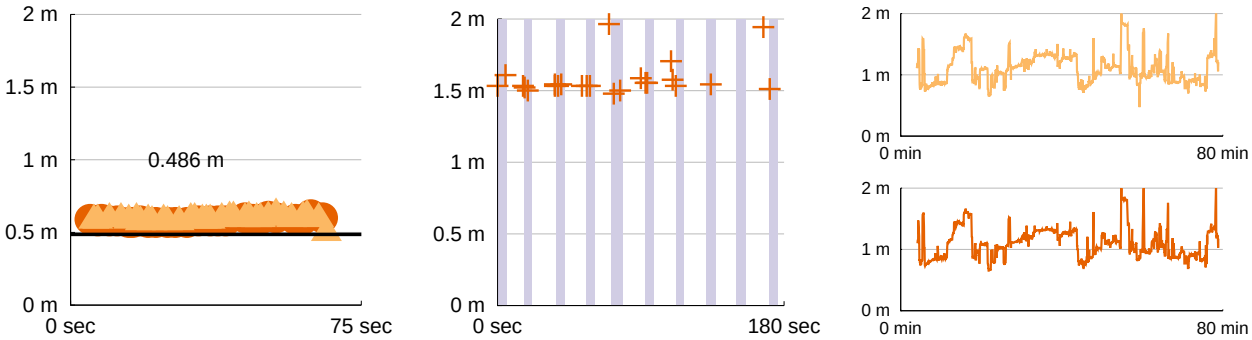
We evaluate Opo's expected group interaction performance with a four-way interaction test (Figure 2.7). We find that sensors do not interfere with each other, with median range measurements falling within 3% of the true distance. Sensors report an average 1.5 s temporal fidelity and 82% packet reception rate. Performance is consistent among the four sensors, with packet reception rates between 77% and 86% and average temporal fidelities between 1.48 and 1.59 s.

Effects of Density

One concern is that sensor density can cause false range measurements through colliding Opo transmissions. We expect these errors to manifest as ranging "spikes", which would increase the spread of measured distances and decrease ranging accuracy. To evaluate the effects of density on ranging accuracy, we placed 18 sensors facing inwards flush along the sides of a 0.38 m x 0.30 m x 0.27 m enclosed box for around 12 hr. Ranging accuracy appears to be in line with our microbenchmarks, although we only measured a sampling of the possible sensor pairings. Opo detected 143 of the possible 153 pairings. In all but nine pairings, 90% of the data readings are within .10 m of each other. In 87% of the pairings, 90% of the range measurements are within .05 m of each other. While it is possible that density introduces consistent, systemic error that our sample hand measurements missed, we find this to be highly unlikely. This test shows that even in dense situations, Opo will likely maintain high ranging accuracy. However, density did affect temporal fidelity. Opos in this scenario report a 8.4 s temporal fidelity. This is expected, since in the current Opo implementation, sensors do not take into account the time they spend receiving, writing to flash, etc, when setting their transmit timers. However, this is still a 2x improvement over the 20 s temporal fidelity of currently deployed interaction sensing systems.

Capturing Human Interactions

To evaluate human interaction sensing, we perform a series of short experiments with six participants. Participants are given as much freedom as possible without breaking the overall



(a) Hallway Encounter (b) Short Interactions (c) Extended Close Encounter

Figure 2.8: Various Pairwise Human Interactions: 2.8a shows a typical hallway encounter where two participants walk towards each other in a hallway, stop and chat for 65 s, and then continue on their way. To test Opo's ability to capture short interactions (2.8b), two participants stood on either side of a hallway corner. At regular intervals, one participant would turn the corner, face the other for 5 s, then go back to his original side of the corner. As 2.8b shows, Opo successfully captures the majority of short interactions. 2.8c shows an extended, professional interaction where two participants sat and collaborated on a project for 75 min. The variations in distance are the result of natural movements, and show what proximity sensing systems miss.

flow of the experiments, meaning we only have a rough sense of ground truth distances.

We evaluate temporal fidelity by examining the time between ranging measurements and ranging accuracy through cross-validation of ranging trends. If only one sensor reports a significant change in range, it is likely a false measurement.

Figure 2.8a: Hallway Conversation

Our first test is a simple hallway conversation to establish that the ranging accuracy from our microbenchmarks carries over to real face-to-face interactions. Two participants walk towards each other, converse for 65 s, and then part ways. Opo sensors report an interaction time of 60 s, a packet reception rate of 98%, and an average of 1.07 s between packets. This matches our ideal temporal fidelity of 1.0 s for a transmit period of 2 s. We find an average range measurement of 0.525 m with a standard deviation of .018 m. True range is roughly 0.49 m, leaving us with a 0.035 m error.

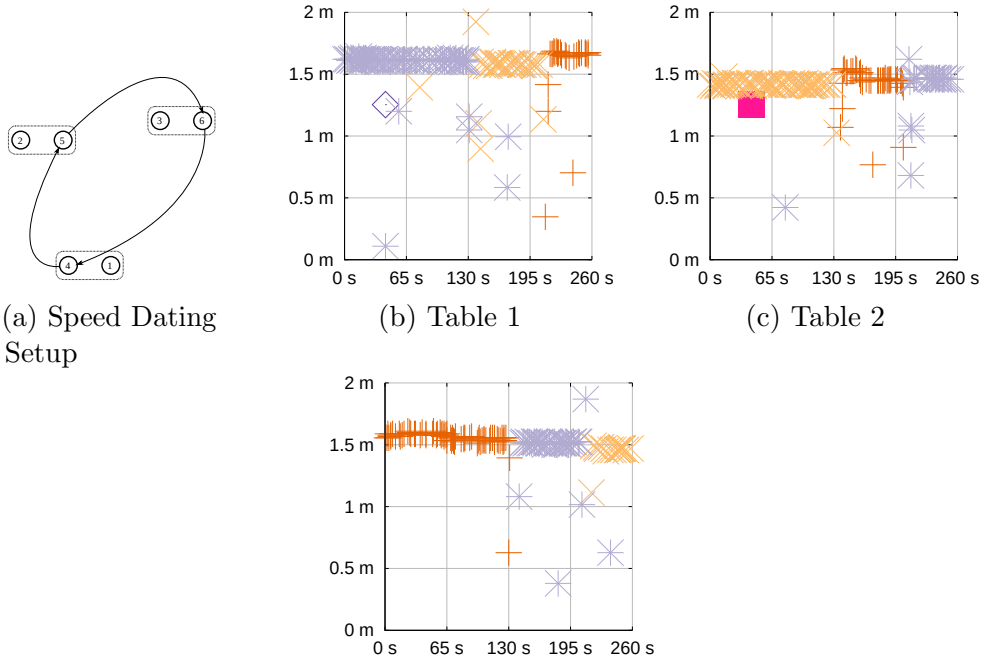
Figure 2.8b: Short Interactions

To evaluate Opo's ability to capture short, spurious interactions, we perform ten 5 s interactions. To do so, we have two participants stand on each side of a hallway corner, with one periodically turning the corner to interact for 5 s, then returning to his side of the corner. We successfully capture 9 out of the 10 interactions. Our total sensed interaction time is 36 s,

which is in line with expectations. While not perfect, Opo's ability to capture short spurious interactions represents a significant upgrade from the current 20 s temporal fidelity of RF scanning sensors.

Figure 2.8c: Close Encounters

To demonstrate the detail that is lost with lower temporal fidelity and proximity sensing instead of range measurements, two participants collaborate on a project for 75 min. Deviations are observed at each sensor, which we attribute to natural movements – leaning in to discuss details, reclining in chairs, and other common actions. This test exemplifies the type of spatial and temporal detail that Opo is capable of providing over state-of-the-art infrastructure-free systems.



(a) Speed Dating Setup (b) Table 1 (c) Table 2 (d) Table 3

Figure 2.9: Speed Dating: 3 tables are laid out, with 1 stationary participant sitting at each table, and 3 participants who rotate between tables (2.9a). Rotation times decrease from roughly 120 to 60 to 45 s. 2.9b, 2.9c, 2.9d represent the interactions at each table, with each rotating participant being represented by a different color and marker. Ground truth is roughly set to 1.5 m, but due to chair position and posture differences, exact error is difficult to determine. However, ranging is precise within interactions, and only 7 of 396 range measurements are unrealistic. Distance variations between interactions are the result of people getting in and out of table chairs.

Figure 2.9: Speed Dating

In spaces such as study rooms and coffee shops, there may be multiple pairwise face-to-face interactions within the same general area. To evaluate Opo's ability to characterize individual interactions within the same general space, we perform a speed dating experiment. We set up three speed dating tables, with three stationary participants and three participants rotating between tables. The duration shortens with each rotation, going from roughly 120 to 60 to 45 s.

We observe an average temporal fidelity of 1.66 s and packet reception rate of 90%, which is in line with our hallway and short interaction tests. Seven out of the nine "dates" are sensed to within 5 s of their true duration. However, two of the 120 second dates are sensed as 108 and 100 seconds. In one of these dates, we found that one participant held his hands in front of his sensor for part of the interaction, causing a gap in range measurements. We speculate that the other interaction error is from a similar cause.

It is difficult to definitively pinpoint false ranges. As people get in and out of seats, we see spikes in interaction distance, which is expected. Other spikes may be caused by people leaning forward or rolling their seats. However, 7 of the 396 range measurements are unrealistic.

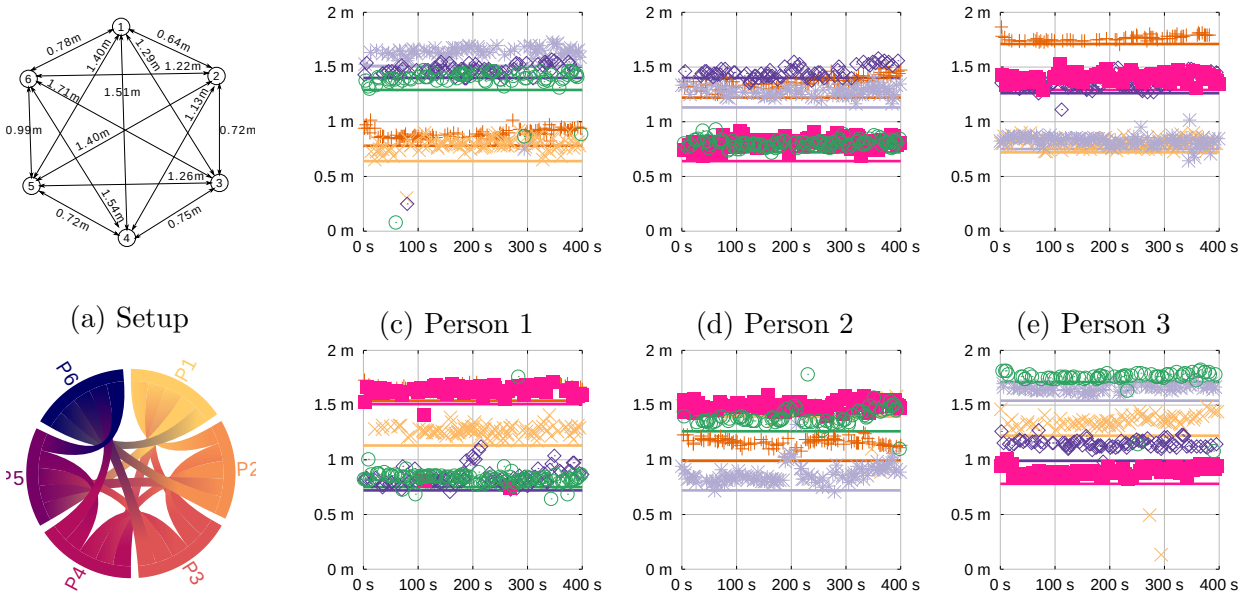
Figure 2.10: Group Chat

We evaluate Opo's ability to capture real group interactions with a 6.5 min six person group chat. Figure 2.10b shows that Opo accurately senses that six people are all interacting with each other for the same amount of time. Each colored arc represents a different participant, with chords representing interactions between connecting participants. Chord width is determined by the total sensed interaction time between two connecting participants.

Spatially, there is a 0.07 m standard deviation in pairwise range measurements. Based on empirical observations, spatial deviations over 0.5 m are likely incorrect range measurements. From a total of 2683 range measurements, we find 14 false ranges. Temporally, we see an average 2.3 s temporal fidelity among pairs of sensors and an average packet reception rate of 80%. Lower packet reception rates are unsurprising since we observe that in group situations people will sometimes turn towards a specific person. As discussed in Section 2.6, temporal fidelity falls with density.

Energy Costs and Model

To evaluate expected battery life, we first calculate the energy cost of each Opo operation. We use these costs to inform a basic energy model from which we extrapolate lifetime estimates. We have not previously discussed data offloading as it is not central to the design of Opo. However, we would be negligent to omit data offloading from our energy model. In this analysis, we assume batch offloading and only burden Opo sensors with storing measurements to flash.



(a) Setup (c) Person 1 (d) Person 2 (e) Person 3

(b) Interaction Times (f) Person 4 (g) Person 5 (h) Person 6

Figure 2.10: Group Chat: 6 participants stand in a circle and chat for around 6.5 min. The setup and rough ground truth is shown in 9a, although participants naturally moved and shifted during the course of the experiment. In 9b, chords represent interactions between participants, with chord width being based on measured interaction time. 9c-9h show the range measurements from each participant/sensor. Rough ground truth is shown as solid lines. 14 of the 2683 measurements are found to be unrealistic.

Energy Costs

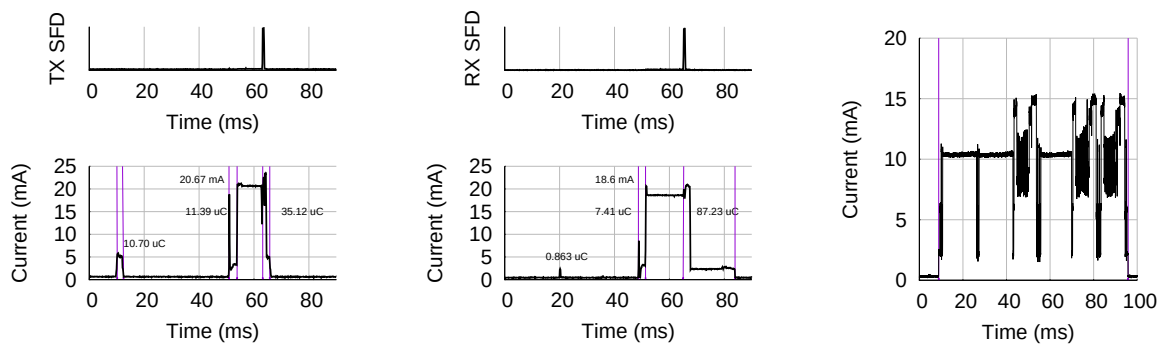
Figure 2.11 captures power traces of each of Opo's fundamental operations. We examine these in detail here.

Opo Ranging Transmission:

$$\begin{array}{rlcl} & E_{\text{UL_PULSE}} & = & 35.31\ \mu J \\ + & E_{\text{TDoA_PULSES}} & = & 784.7\ \mu J \\ \hline & E_{\text{TX}} & = & 820\ \mu J \end{array}$$

Due to software inefficiencies, the energy required to send an RF packet is artificially high. The default TinyOS CC2420 driver uses a 9 ms backoff timer—time 53 to 62 in Figure 2.11a—when transmitting a packet, where the radio is in a high power receive state. While this is helpful in synchronized sensor networks, this backoff timer serves no purpose for Opo. We estimate that modifying the driver and getting rid of the backoff timer would save 564 μJ of energy[2], a 72% reduction. Our current generation of Opo uses a different radio with similar power draws, and does not have this back-off timer problem. However, we have not validated

[2] $(18.8\text{ mA} - 426\ \mu A) \times 3.3\text{ V} \times 9.3\text{ ms} = 564\ \mu J$



(a) TX Current Draw (b) RX Current Draw (c) Flash Write Current Draw

Figure 2.11: Current Draw Traces: Red lines represent various components (UL TX frontend, radio, mcu, flash) waking up and performing tasks. In 2.11a, the high current draw time between 53-62 s is due to a radio driver inefficiency. The ranging reception shown in 2.11b is done at 6 m to increase the visibility of the high powered, radio reception state. The SFD graphs for 2.11a and 2.11b show when the relevant radio transmission and reception actually occur. The smaller current spikes occurring before 50 s in 2.11a and 2.11b are from the UL frontends and MCU waking up, while the larger current spikes are from the radio turning on.

the power draw of the current generation of Opo to the same degree as the version used here, so we conservatively keep these flaws in our power calculations

Opo Ranging Reception:

$$\begin{array}{rrcl} & E_{\text{MCU_UL_INTERRUPT}} & = & 2.85\ \mu J \\ + & E_{\text{RF_IDLE_LISTEN}} & = & 61.4\ \text{mW}\times(range)\ \text{ms} \\ + & E_{\text{RF_PACKET}} & = & 184.4\ \mu J \\ + & E_{\text{PROCESSING}} & = & 127.9\ \mu J \\ \hline & E_{\text{RX_RANGE}} & = & 315.2\ \mu J + \text{E}_{\text{RF_IDLE_LISTEN}} \end{array}$$

The energy consumed idly listening by the RF frontend will vary as a function of the distance between the transmitter and receiver. Figure 2.11b shows a 6 m range to exaggerate this effect and clearly identify the RX idle listening time.

Write Batch of Range Data to Flash: $E_{\text{WRITE_FLASH}} = 2.83\ \text{mJ}$

Our Opo implementation writes flash in 36-sample[3] batches. Thus, to get the actual cost of a range reception:

$$E_{\text{RX}} = E_{\text{RX_RANGE}} + \frac{E_{\text{WRITE_FLASH}}}{36} = 393.8\ \breve{\text{J}} + E_{\text{RF_IDLE_LISTEN}}$$

Static Power While Idle:

$$P_{\text{IDLE}} = 148.5\ \mu W$$

In idle mode, Opo is in its lowest-power state. The radio is powered off, the MCU is in deep, RAM-preserving sleep, and only the ultrasonic wakeup circuit is powered.

[3] $512\ \frac{\text{bytes}}{\text{page}} / 14\ \frac{\text{bytes}}{\text{record}} = 36$ records per page.

Energy Model

To reason about energy, we make a simplifying assumption that there is no packet loss. Our analysis double-counts Opo's static power draw as baseline power was not removed from individual energy costs. As the idle power is $10 - 20\times$ lower than the active power and Opo's duty cycle is relatively low, the impact on our model is not significant.

We begin with the simplest case: two sensors facing one another. Every two seconds, each sensor will perform one ranging transmission and one ranging reception. We can express the energy each sensor consumes as a function of time t and average distance d_{avg} between the sensors:

$$E_n(t, d_{avg}) = t \times 148.8\ \mu W + t \times \frac{1 \text{ range}}{2 \text{ sec}} \times$$
$$(820\ \mu J + 393.8\ \mu J + 315.2\ \mu J + 61.4 \text{ mW} \times d_{avg})$$

Next, we extend our model to include multiple sensors. Our aim is to place an upper bound on energy consumption. Opo consumes the most energy when it sees the greatest number of sensors (more ranging receptions). Our model takes this to its natural extreme, $n + 1$ sensors arranged such that all $n + 1$ sensors are facing one another (e.g. in a circle):

$$E_n(t, d_{avg}, n) = t \times 148.8\ \mu W + t \times \frac{1 \text{ range}}{2 \text{ sec}} \times$$
$$(820\ \mu J + n \times (393.8\ \mu J + 315.2\ \mu J + 61.4 \text{ mW} \times d_{avg}))$$

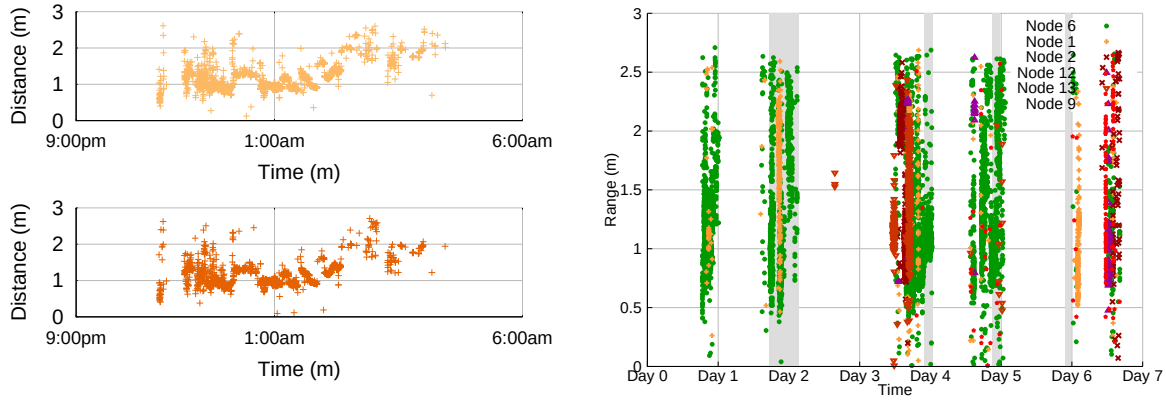
Lifetime Estimation

We consider a hypothetical deployment with an average of 5 visible sensors, 2 meters (6 ms RX) average range, and 8 hours of interaction per day. In this setting, each Opo sensor would consume about 126 J of energy per day. A week-long deployment would require a 66 mAh battery.

Real-World Pilot Deployment

To validate Opo's real-world operation, we deployed Opo sensors among eight participants for a week. This deployment produced 47,189 range measurements. Sensors had an average battery life of 93 hours, or 11.6 business days, on a 40 mAh battery the size of a dime. Sensors used an average of 137 J a day, compared with the 126 J from our model.

Establishing ground truth proved challenging. Not all participants were comfortable with video taped ground truth, and efforts to log all interactions in journals quickly proved untenable, even among just eight people. Generally speaking, we see clustering of ranges and cross validation of ranging estimates similar to that seen in our extended close encounter



(a) Two participants partying

(b) Sample data from one participant.

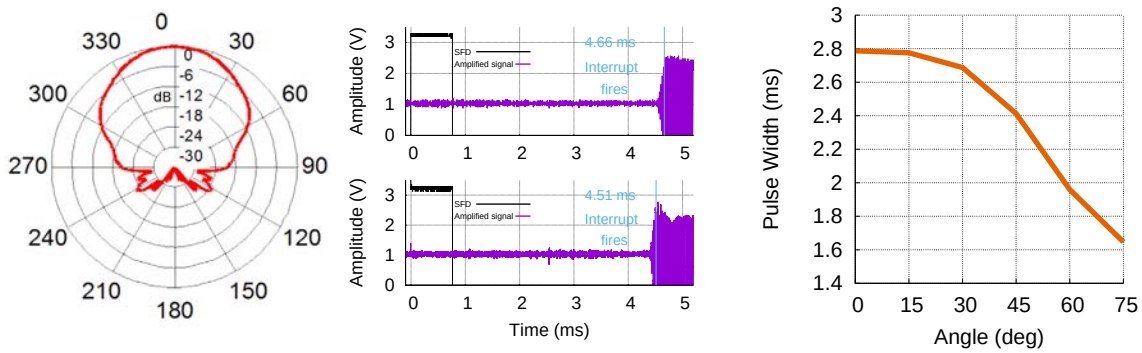
Figure 2.12: Real-World Deployment: 2.12a shows Opo measurements from two participants during a party. Measurements from one participant's sensor is consistently validated by measurements from the other participant's sensor. 2.12b shows sample data from one participant. The shaded areas are time spent by the participant in uninstrumentable spaces (e.g. restaurants).

experiment (Figure 2.8c). A 9 hour slice of time where two participants partied together is shown in Figure 2.12a as an example of this clustering and cross validation. We find a median 5 s temporal fidelity, as opposed to the 1 s ideal, which is still a 4x improvement over current interaction sensing systems. Participants wore the sensors using both lanyards and magnetic clips, and reported no discomfort.

Even in this short deployment we see the importance of infrastructure-free operation. Figure 2.12b shows the results from one participant, who spent 2 days entirely outside of academic settings, which is shown as gray shaded areas. Even on days where participants were in possibly instrumentable buildings during the day, significant time was spent in other uninstrumentable areas, such as restaurants during dinner, bars after work, etc.

2.7 Reducing Angular Offset Errors

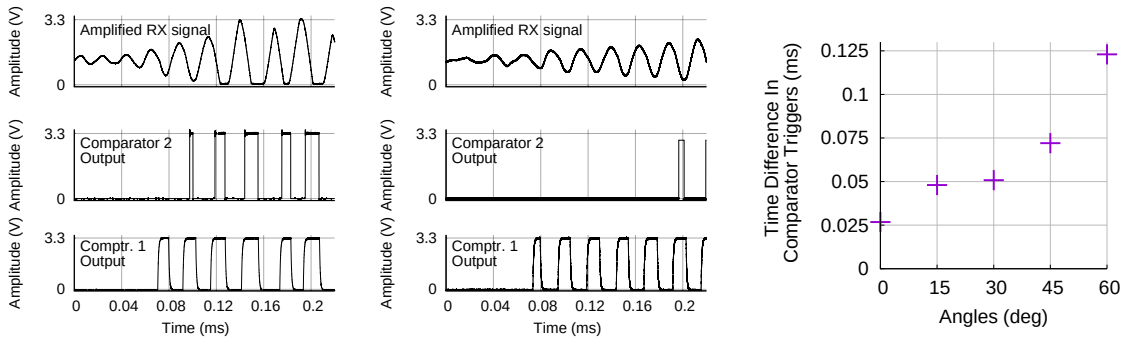
Finally, we would like to discuss the angular ranging error shown in Section 2.6. In short, angular offsets between sensors, like those among adjacent people standing in a circle, increase ranging error. This occurs because the ultrasonic transducer's transmit power and receive sensitivity decrease with offset, as shown in Figure 2.13a. This results in a delayed received ultrasonic rise time, leading to a delayed timing capture, which is interpreted as a longer range. Figure 2.13b illustrates this delay by comparing the output of the amplified ultrasonic signal received under face-to-face (bottom) and high angular offset ranging (top). The angular offset causes a 150 μs delay, which results in a 5 cm estimation error at room temperature. The reduced signal strength also results in a shorter pulse length. Figure 2.13c shows that pulse



(a) Ultrasonic Beam Angle (b) Angular (top) vs Direct (bottom) Ranging (c) Received Pulse Width vs Angle

Figure 2.13: Effects of Angle of Arrival.

length decreases with with an increase in angular offset. We hypothesize that the reduced power at angular offsets manifests as a longer time for the transducer to reach resonance.



(a) Face to Face Ranging (b) Angled Ranging (c) Measuring Rising Edge Slope with Comparators

Figure 2.14: Exploring Ultrasonic Rise Time

Measuring pulse duration in heavy multipath settings may be challenging as reflections can extend the received pulse duration. An alternate approach to detecting angular offsets could be to estimate the slope of the envelope of the received ultrasonic signal using two comparators with different voltage thresholds. Figure 2.14a shows the received signal and the output of two comparators with two different thresholds for a head-on range. Figure 2.14b shows the same signals for a 60° offset angle range, simulating six people standing in a circle. Our limited testing (Figure 2.14c) shows an inverse relationship between comparator trigger time differentials and angular offset. This suggests that future UL/RF TDoA systems could use this technique to reduce offset angle induced error sand possibly estimate interaction orientations.

2.8 Summary

Face-to-face interactions are important in many settings, but unobtrusively and efficiently capturing them for study has remained notoriously difficult. In this paper, we present a low-power system that can sense face-to-face interaction distance in an infrastructure-free manner. The key element that enables this work is a novel ultrasonic wakeup frontend that can be built from commodity components. With very low power draw, sensors can be miniaturized to the point that they are the size of a large lapel pin, yet keep a 4-day lifetime with a 40 mAh battery.

This enables us to keep power-hungry radios mostly off without resorting to duty-cycled neighbor discovery protocols that sacrifice discovery latency to achieve low power. The power and bandwidth requirements of our combined ultrasonic/radio broadcast ranging scale linearly with the number of co-located sensors, whereas pairwise approaches like two-way time-of-flight ranging scale quadratically. As a sensor, Opo shows that ultrasonic wakeup is viable using commodity components, and that we can utilize this to build low-power, high-resolution, infrastructure free interaction sensors, which two-decades of prior work did not achieve. In addition to using Opo to sense face-to-face interactions, we build upon the Opo ranging primitive to also sense hand-washing behavior in Chapter 4.

However, to us, Opo also represents a different way to think about making complex wearable sensors. Instead of trying to improve how efficiently wearable sensors actively perform tasks, the power constraints that limit complex wearable sensing applications may be solvable by finding a low-power, passive method of detecting phenomenon of interest. Opo was first published in 2014. In 2020, Socitrack represents significant advancements in every part of the traditional active interaction sensing architecture, and uses significantly newer microprocessors and radios than Opo [41]. Even so, it draws 40 x the power of Opo. Although Socitrack has advantages over Opo in other applications, for the purpose of better understanding disease-spread, we believe Opo still represents the state-of-the-art. We believe this shows the power of this idea, and we utilize this way of thinking in Chapter 5 to design a novel cough sensor that does not require users to constantly record their audio environment. However, before we get to these chapters, we first take a deeper dive into evaluating the benefits and limitations of using Opo.

Chapter 3

Deployment Experiences and Evaluation

3.1 Motivation and Background

Our evaluation of Opo shows that Opo provides high-resolution interaction distance and time without compromising deployability. However, this does not necessarily mean that the data Opo provides will provide significant value over currently used lower-resolution sensors to better understand disease-spread. If Opo's high-resolution data can largely be inferred from BLE proximity sensing, then Opo may not provide value beyond a few studies to properly calibrate BLE sensing. Although we heavily evaluated Opo's core functionality, we also had questions of how flexible Opo is in real-world scenarios.

To better understand these problems, we explore two questions. First, how does the data from Opo differ from network graphs generated by other sensors, in particular Bluetooth Low Energy proximity sensing, the current default sensor used to collect contact data for disease spread? Second, how well do Opo and other chest-worn sensors capture face-to-face interactions in environments that can restrict an occupant's ability to fully their turn their body towards other occupants?

3.2 Opo vs BLE Proximity Sensing

To explore the real-world differences between networks generated by Opo and lower-resolution sensors, we deploy Opos and a commercial BLE proximity sensing solution in a Midwest Office Deployment. This allows us to examine both the individual and network level differences between data from Opo and typical low-resolution sensor deployments. While it is intuitive that Opo generates much higher resolution data in both the time and distance domains than BLE sensing, what is less intuitive is what these differences mean when examining a network of individuals. For example, if a group of individuals typically interacts at the same distance from each other and exhibit uniform random mixing, than it is possible that at the

network level that Opo data is not much more informative than lower-resolution data. Our office deployment allows us to explore both quantitatively and qualitatively the benefits and drawbacks of using Opo compared to traditional BLE sensing in a real-world non-academic setting.

3.3 Study Design: Midwest Office

	N / Median	**% / IQR**
Age	46	39,55
Gender		
Male	13	57%
Female	10	43%
Race/Ethnicity		
White, non-Hispanic	20	87%
Other	3	13%
Currently Married	15	65%
Education Level		
Some college or less	3	13%
College graduate or more	20	87%
Office Type		
Individual	12	52%
Shared	2	9%
Cubicle	9	39%
Supervisor	7	30%

Table 3.1: Office Site A Demographics

We conducted a pilot longitudinal network study of contact patterns and hand hygiene in two Midwestern offices (designated as site A and site B hereafter) in the United States during March 2018. The office worksites were chosen based on convenience but comprise typical layouts of many office workplaces in the US; including open-air cubicles, private offices, bathrooms, and shared breakroom space. Participants were recruited via email through the company listserv, posters in common areas, and in-person by study staff. To be eligible, employees needed to be at least 18 years old and plan to work in the office for at least part of the follow-up period. At each office site, individuals were enrolled and followed during working hours for one work-week (Monday 10am to Friday 3pm) during consecutive weeks during March 2018. At enrollment, participants provided informed consent, completed an online baseline survey, and issued two sensors for the study period. At the end of the study period, participants responded to an exit survey and were given a $25 gift card as an incentive for participation. We conducted this pilot in collaboration with epidemiologists at the University

of North Carolina. This data collection and research were approved by the University of North Carolina Institutional Review Board and all participants provided written consent.

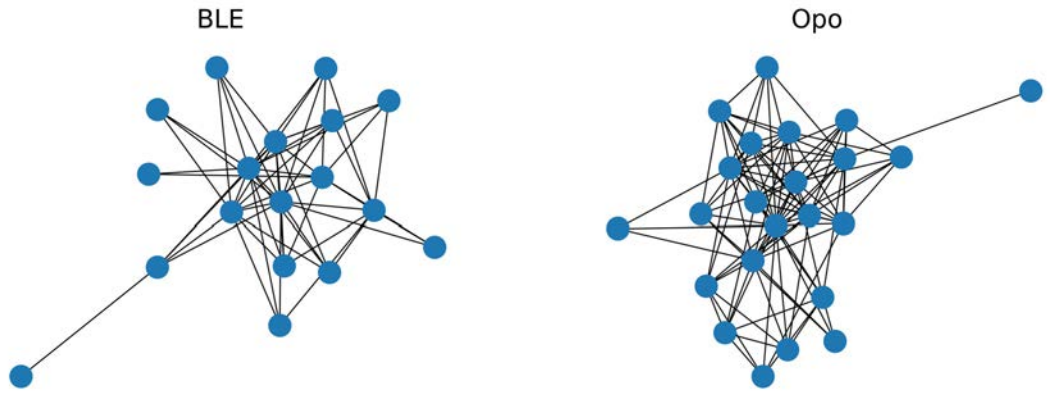
Participants were given Opo sensors and those with compatible smartphones were given a Bluetooth Low Energy beacon and asked to install a phone app from Ethica (https://ethicadata.com/). Those who installed the phone app also received a Bluetooth Low Energy beacon. The Ethica phone app periodically listened for BLE broadcasts every 5 min from the beacons, while the Opo sensors performed a ranging transmission on average every 2 s. The Ethica phone app would typically capture multiple multiple packets from a nearby beacon during each listening period, so we filtered out duplicate packets in the same listening period for our analysis.

We focus our analysis on office site A. In office site B, we found that the lighting control systems in parts of the office operated at the same ultrasonic frequency as Opo, 40 kHz. This resulted in them interfering with Opo's operation, and resulted in large chunks of missing data. We could not find any data on how commonly lighting-control or motion-detection systems utilize 40 kHz ultrasonic signals, but we note that we have used Opo in a variety of buildings across three states, and this was the only instance where we found significant interference. We recruited 23 participants at office site A and their demographics are shown in Table 3.1. The 23 participants were provided with Opos, and 18 of them also used the Ethica BLE phone app and beacon.

3.4 Office Study Evaluation

All 23 participants registered at least one Opo interaction with another participant, and all 18 participants with the Ethica phone app registered at least one BLE interaction with another participant. To examine network level differences in data from the two sensors, we generated un-weighted graphs from the two data sets using NetworkX[72], as shown in Section 3.4. Nodes in the graphs represent participants, and edges represent interactions. In the BLE network, an edge between two participants is established if at least one participant's Ethica phone app received a packet from the other participant's BLE beacon. In the Opo network, a edge between two participants is established if at least one successful Opo ranging event occurred between the participants' Opos. We then used NetworkX to calculate various network statistics. During our analysis, we found that many of our variables had non-linear relationships, so we examine rank correlation here instead of the more common linear correlation.

Table 3.2 shows an overview of some network characteristics and interaction times generated from the Opo and Ethica BLE data. We selected average degree centrality and closeness centrality as sample network statistics because they are often used to identify and determine the effects of super spreaders in disease spread. Centrality in general measures how important, or central, a node is to a network. Degree centrality is the proportion of other nodes a node in a network is connected to, while closeness centrality is a measure of how close a node is to every other node in a network. Specifically, the closeness centrality of a node a is the



(a) BLE Contact Network (b) Opo Contact network

Figure 3.1: Contact networks for office site A. Each node represents a participant, and a connection between two nodes indicates that those two participants had at least one interaction using that sensor. Both sensors produce similarly connected graphs. However, further examination of the data indicates that this is somewhat coincidental. The BLE sensor misses many short-duration interactions which are critical to connectivity in the Opo network, but makes up for this by falsely detecting participants in adjacent offices and cubicles as interacting.

reciprocal of the average shortest path distance to a over all other nodes in the network. In general, for any centrality measure, the more interconnected a network is the higher the average centrality will be.

On the surface, the overall network characteristics of the Opo and BLE data look the same while the interaction times sensed for our participants is wildly different. Excluding participants that only have an Opo does not significantly change this. However, while the overall network structure between the Opo and BLE data is similar, we found significant differences in terms of how individual participants appeared in the network. That is, some participants were far more connected and central to networks in one set of data than the other. Ranking participants with both Opo and Ethica BLE sensors by centrality, we found that the Spearman's rank correlation coefficient for both degree and closeness centrality generated from the Opo and BLE data sets is less than 0.25. This indicates that how central a participant based on one sensor technology is only weakly correlated with how central they are with the other sensor technology.

The key to why this occurs can be found in examining the interaction times of the participants using Opo verses the Ethica BLE solution. To calculate interaction times for Opo, we simply multiplied the number of Opo packets involving a participant by 2 s. Similarly, we multiplied each Ethica packet received by a participant's phone app by 5 min to calculate interaction times for the BLE data. While this method slightly underestimates interaction times because it does not take into account missing packets, it gives us a good approximation

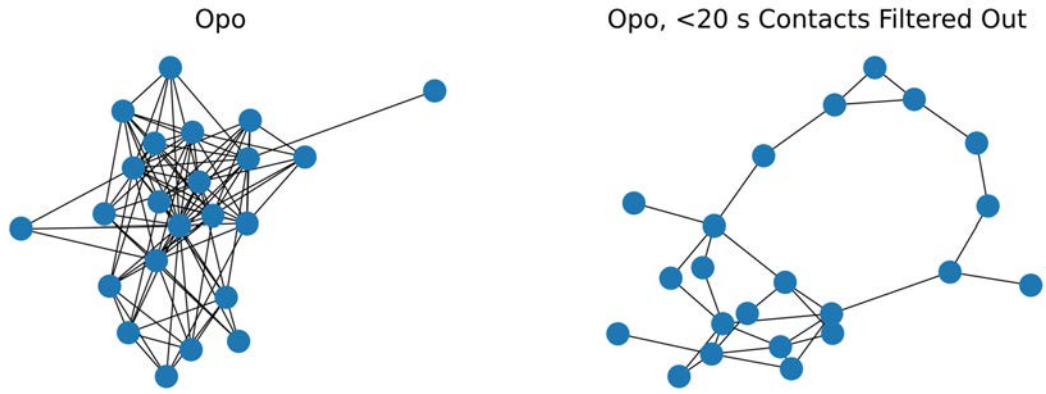
	Mean	Median	IQR
Opo			
Degree Centrality	0.41	0.36	0.27, 0.55
Closeness Centrality	0.63	0.61	0.57, 0.69
Interaction Time	1.20 hr	0.77 hr	0.07, 1.23 hr
BLE			
Degree Centrality	0.37	0.29	0.24, 0.53
Closeness Centrality	0.61	0.57	0.55, 0.65
Interaction Time	13.94 hr	8.13 hr	3.19, 21.73 hr

Table 3.2: Comparison of centrality and interaction time between Opo and BLE generated networks. As mentioned in Section 3.4, the similar centralities, or measures of how connected a network is, between Opo and BLE is somewhat coincidental. Although we do not have ground truth for the interaction times, based on manually examining the data, our observations during the study, and anecdotal discussions with participants, suggest the Opo interaction times are more accurate. The BLE interaction times are most likely the result of participants in adjacent work spaces being falsely detected as interacting.

of what interaction times are like for our participants. As Table 3.2 shows, the interaction times from the BLE data is an order magnitude higher than the interaction times from Opo data. In addition, the relative amounts people interacted with each other was also wildly different based on the sensor technology used. Ranking participants with both sensors by their total interaction time, the Spearman's rank correlation coefficient for Opo and BLE interaction times is -0.18. That is, having a higher interaction time rank using one sensor technology was weakly correlated with having a lower interaction time rank using the other sensor technology. Based on our observations during the studies and anecdotal conversations with participants, we believe the reason for this is that even though the majority of participants worked in individual offices or cubicles, many of them were located in close proximity to each other. This means that even when participants were just sitting in their office or cubicle, the BLE sensor would often report that they were interacting with another participant in a nearby office or cubicle. In the case of the two participants who were in a shared work space, we manually examined the data and found that the BLE sensor does indeed detect them as almost constantly interacting.

Time Resolution

In addition to comparing Opo's network structure and interaction time with BLE data, we seek to better understand what benefits having high resolution interaction time and distance data has. While the 5 min time resolution of the Ethica sensor system is typical of solutions involving phones, dedicated infrastructure-free wearables such as the WREN [73] often have a 20 s time resolution.



(a) Opo Contact Network

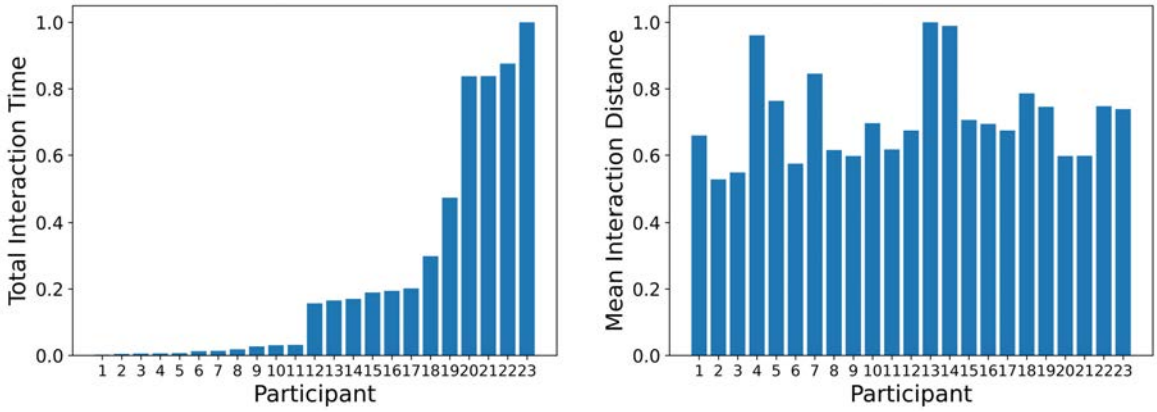
(b) Opo Contact Network, No Short Interactions

Figure 3.2: The importance of short duration interactions. Here we show the Opo generated contact network with all interactions and with interactions less than 20 s filtered out. Filtering out short interactions only results has a 16% reduction in total interaction time, but results in a significantly less connected contact network.

To explore the effects of having a 20 s interaction time resolution, we generate an Opo contact network with interactions ¡ 20 s filtered out and compare it with a full Opo contact network, as shown in Section 3.4. We calculate individual interaction times by separating out Opo packets between two participants into segments where each packet is within 7 s of the previous packet in the segment. While in theory Opo packets in an interaction segment should only be separated by 2-3 s, or the transmission rate of Opo, in practice we have found that sometimes Opos miss a packet or two for a variety of reasons. Removing all interactions less than 20 s reduces the total interaction time of our participants by 16%. More significantly, removing these short interactions significantly reduces the connectivity of our network, as shown in Figure 3.2b. Quantitatively, removing interactions less than 20 s reduces our mean degree centrality from 0.41 to 0.14, and our mean closeness centrality from 0.63 to 0.34. This matches our general intuition for our everyday workplace interactions: while we talk to some people a lot, others we normally just bump into in the hallway. While these hallway interactions do not make up the bulk of our overall interaction time, they significantly increase the connectivity of our contact networks. Even though these interactions only account for 16% of our total interaction time, missing them could significantly complicate contact tracing, modeling disease spread, and identifying super spreaders.

Distance vs Proximity

Typically, contact tracing sensors used to inform disease spread use proximity sensing techniques, such as the BLE proximity sensing solution of the Ethica app. Even in fairly controlled



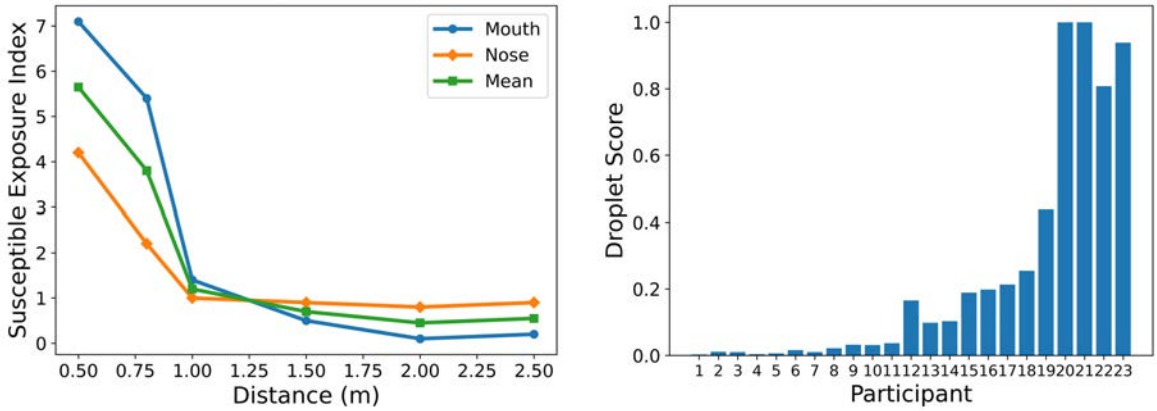
(a) Participants, ranked by interaction time (b) Participants, ranked by interaction distance

Figure 3.3: Participants ranked by normalized total interaction time and mean interaction distance. Total interaction time is only weakly correlated with mean interaction distance, with a Spearman's rank correlation coefficient of 0.16.

scenarios, these techniques typically provide a rough sense of if users are within a 2-3 m proximity rather than a true distance estimate. However, as discussed in Section 2.2, we know that the relationship between interaction distance and disease spread is more complicated than a binary proximity.Intuitively, having distance-measurements rather than proximity would allow researchers to better study and understand the relationship between interaction distance and disease spread, and better model how many diseases spread in the real world. However, if interaction time can serve as a good enough proxy of interaction distance or if differences in interaction time is the dominant factor in disease spread, then perhaps high-resolution distance is not an important feature for understanding or modeling disease spread.

To explore this, we first check if interaction time is a good proxy for interaction distance. Figure 3.3a shows the study participants ranked by total interaction time, while Figure 3.3b shows the participants' mean interaction distances. For easy comparison, we show normalized interaction times and distances here. As we can see, total interaction time and mean interaction distance do not correlate very well, and the Spearman's rank correlation coefficient between the two variables is only 0.16.

Next, we explore how important interaction distance is from a disease spread perspective. How interaction distance directly relates to infection rate is an ongoing area of research, and we have not seen a model for this for any disease. Instead, we incorporate prior work on how droplet exposure relates to interaction distance. Droplets, or tiny drops of water commonly expelled during breathing, coughing, and sneezing, are thought to be an important transmission vector in many diseases, including COVID-19 [12]. Liu et al explore the relationship between droplet exposure and interaction distance using mannequins with artificial lungs and a heavily controlled 4.2 m x 3.2 m x 2.7 m room [67]. As part of their



(a) Susceptible Exposure Index vs Distance

(b) Participant Droplet Scores

Figure 3.4: Calculating droplet exposure scores for participants. Figure 3.4a shows a simplified relationship between between interaction distance and droplet exposure [67]. We use this to calculate how many droplets each participant exposed other participants to during this study, and show normalized scores in Figure 3.4b. We find that droplet scores only correlate moderately (0.30 rank correlation) with interaction time, meaning that high-resolution interaction distance may provide significant value in better understanding disease spread.

analysis, Liu et al show the relationship between susceptible exposure index, which is a measure of how much of another person's droplets one is exposed to, and interaction distance. If person A and person B are interacting in a room, person A's susceptible exposure index is the amount of person B's droplets they are exposed to relative to if person B's droplets were evenly spread throughout the room. Liu et al provide averaged exposure indexes for a variety of interaction distances across 21 indoor air conditions for both mouth and nose breathing. We show these and an average between nose and mouth breathing in Figure 3.4a.

Using the averaged mouth and nose breathing susceptible exposure indexes, we calculate an exposure index for each participant based on their mean interaction distance. We then use this to calculate a droplet score for each participant, where

$$droplet\ score = total\ interaction\ time * exposure\ index \tag{3.1}$$

Figure 3.4b shows a normalized chart of droplet scores for participants. Droplet scores do correlate with interaction time better than mean interaction distance does, but the Spearman's rank correlation coefficient between droplet scores and total interaction time is still only 0.30. The exact relationship between droplet exposure and infection rate is unclear. But, this result shows that information on interaction distance rather than binary proximity could be significant in understanding disease spread. We note that droplet scores and interaction time appear to have a linear relationship. However, we believe this is just an artifact from of our simplified model in Figure 3.4a, which assumes that the relationship is composed of linear

segments. Based on Liu et al's paper, the real relationship between susceptible exposure index and interaction distance is likely far more complex and non-linear [67].

Overall, our data shows that BLE data can drastically over-estimate interaction times and distort interaction time distributions. For applications where it makes sense to weight network edges by interaction time, such as disease-spread, this can result in significant differences when using Opo and BLE sensors. Furthermore, our analysis shows that high-resolution interaction distance is difficult to infer from low-resolution sensors, and that this information may prove valuable in better understanding and forecasting disease-spread.

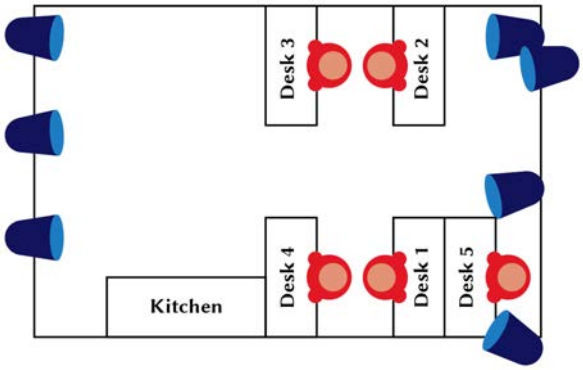
3.5 Exploring Chest Worn Sensors

Chest worn, directional contact sensors such as Opo measure the distance between two people when their chests face each other. This type of data falls under the field of body proxemics, which is informative for a variety of applications [74]. However, for applications such as modeling disease spread, researchers are more interested in when peoples' heads are turned towards each other. To the best of our knowledge, no interaction sensor directly measures this. Sensors that rely on radio proximity sensing or ranging assume that omni-directional distance or proximity can serve as a good proxy for face-to-face interactions, while directional sensors such as Opo assume that when people interact, their chests typically face the same direction as their heads.

Intuitively, we expect people to prefer keeping their bodies and heads oriented in the same direction when possible rather than straining their necks. However, it is obvious that sometimes people simply turn their heads to interact. In particular, we speculate in shared work environments, furniture might restrict peoples' abilities to conveniently orient their bodies and heads in the same direction when interacting. Furthermore, we note that in many work environments, it is common for one person to face another person when interacting, but for the second person to be facing a computer screen or other object during the interaction. We refer to face-to-face interactions where both people are facing each other as two-sided, and face-to-face interactions in which only one participant is facing the the other as one-sided. To explore how significant these problems are to chest-worn sensors, we conducted a pilot study using a motion capture system and Opo in a shared computer lab space.

3.6 Study Design: Motion Capture

In this study, we deploy Opo alongside an Opitrack motion capture system in a shared computer science lab setting during a typical workday [75]. We outfitted the shared lab space with seven Optitrack motion capture cameras along the ceiling and outfitted study participants with motion capture markers and Opos. In addition, we manually observed participants during the study. Each desk had at least one large monitor on it, and participants typically sat in rolling office swivel chairs at their desks while working. The configuration



(a) Work-space Layout

(b) Motion Capture Marker Setup

Figure 3.5: Experimental setup. Figure 3.5a shows the overall layout of our work-space and where our participants sat. The desk number corresponds to the participant ID number used in our charts below. The blue cones indicate the general location of the Optitrack cameras, which were mounted to the ceiling. Figure 3.5b shows how we mounted motion capture markers to our participants.

of the shared space, participants, and motion capture cameras is shown in Figure 3.5a. we note that participants 1 and 4 and participants 2 and 3 are normally within 2 m of each other. Therefore, even a perfectly accurate RF interaction sensor would detect those pairs of participants as constantly interacting. The spacing was such that while sitting, it was often inconvenient for participants to fully turn their chairs and bodies to interact with others in the shared lab space. To track facial orientation and location, participants wore baseball caps outfitted with Optitrack motion capture markers, as shown in Figure 3.5b. Each baseball cap was outfitted with a different geometrical configuration of markers, allowing the motion capture system to track the orientation and location of each participants' head. Finally, participants also wore an Opo sensor on their chests using a lanyard. This experiment lasted for one work day, and participants wore the sensor and motion capture markers only while they were in the shared lab space. Participants were people who normally occupied the lab space and recruited via email. They were instructed to simply behave as they normally do during

the experiment. In total, five participants took part in the experiment. No compensation was provided and the experiment was approved by the University of California, Berkeley IRB committee.

Data Processing

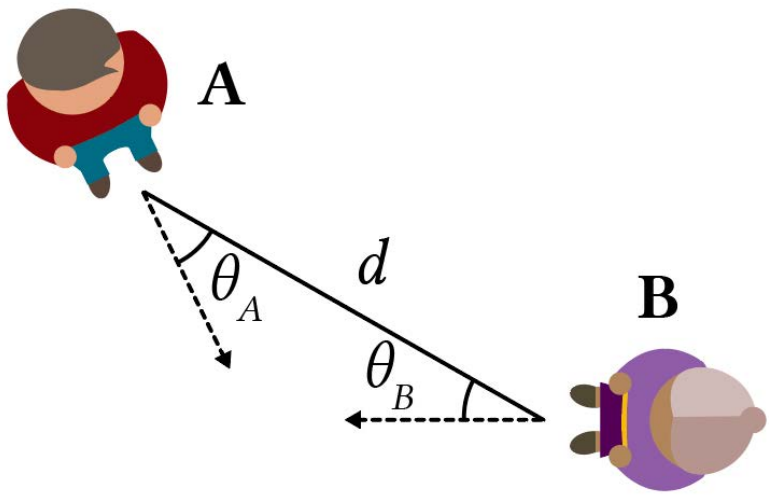


Figure 3.6: Calculating relative angle. Relative angle measures how many degrees a person would have to turn to be directly facing another person. Here we see the the interaction distance d, the relative angle from person B to person A, θ_B, and the relative angle from person A to person B, θ_A. While the interaction distance between persons A and B is symmetrical, the relative angles between them need not be. E.g, if person A has to turn a shorter distance to face person B than person B has to turn to face person A, then the relative angle of A to B (θ_A) will be smaller than the relative angle of B to A (θ_B). An extreme example of this would be if person A was standing behind person B in line. Since person A is facing person B, θ_A would be zero degrees. In contrast, person B would have to turn all the way around to face person A with a θ_B of 180 degrees. If two people are are both directly facing each other, each of them has a relative angle of zero degrees to one another.

The Optitrack system gives us precise location and cardinal orientation for each participant's face at 120 Hz. Using this data, we generate 120 Hz interaction distances and relative angles for each pair of participants, which can be seen in Figure 3.6. Relative angle is simply how many degrees each participant in the pair has to turn to face the other, which we use to determine if one participant is facing another. We then filter out instances where the interaction distance between two participants is over 2 ml, which is the maximum distance Opo works at and is typically the maximum relevant distance for modeling disease spread.

Using the participant locations, we also filter out data points where barriers such as computer screens separate two participants.

To determine if participants are facing one another, we first define a maximum relative angle. Given a pair of participants pA and pB, we say that pA is facing pB if the relative angle of pA to pB is less than the maximum relative angle. Intuitively, we expect the maximum relative angle to be close to zero In practice, we found that this was not true. First, the Optitrack system provides the location of the center of each person's head. This means that our relative angle calculation is specifically how many degrees pA has to turn such that the center of their face is aligned with the center of $pB's$ head. However, we observed that often times when participants faced each other, they did not line up the center of their heads. Second, we observed that the baseball caps that we mount the motion capture markers to sometimes shifted while being worn. Based on our observations and some experimentation, we set a maximum relative angle of 40°. We generate data sets for instances where both participants in a pair are facing each other, for instances where at least one participant is facing the other, and a proximity data set that does not take into account orientation.

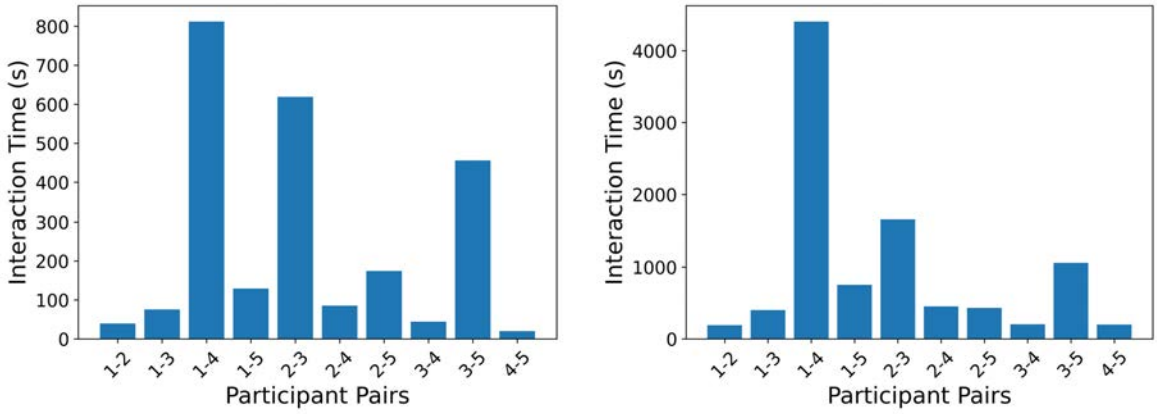
Finally, we segment out these data points into interactions by grouping together data points that are within 10 s of each other. While the Optitrack system provides 120 Hz data, due to monitors and other furniture the system could briefly. lose track of participants. We then calculate total interaction times for each pair of participants by adding up the duration of each individual interaction between the pair. We also use this process to calculate interaction times for the Opo data.

3.7 Preliminary Results: Motion Capture

Due to the limited number of participants and limited study duration, we present our evaluation as preliminary results. We intended to re-run this study with a longer duration and some slight modifications to our motion capture mounting system, but we were not able to update and re-run this study before the COVID-19 pandemic eliminated lab access. Because of these issues, we also limit our correlation statistics to rank correlation even though many of our variables have linear relationships with each other. However, we believe the results here are still informative. Here we examine how often both participants in a pair face each other vs how often only one participant faces the other, and how well our Opo data matches the face-to-face interactions.

Face-to-Face Interactions

As Section 3.7 shows, participants in this lab space would often have interactions in which one participant faced another but the other participant was looking in a different direction. For each participant pair, between 50%-80% of face-to-face interactions involved only one participant faced the other. Observationally, this often occurred when one participant faced a computer screen during an interaction with another participant. In particular, we note



(a) Face-to-Face Interaction Time, Two Sided (b) Face-to-Face Interaction Time, One Sided

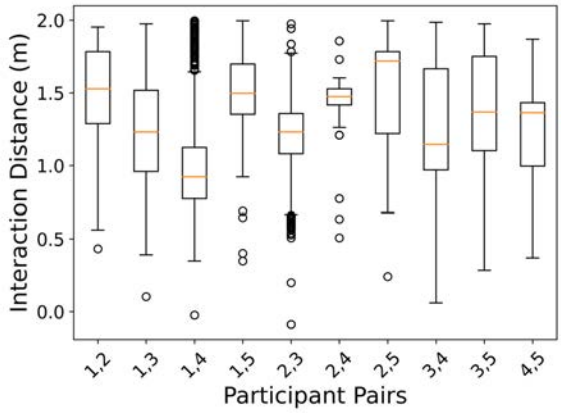
Figure 3.7: Face-to-Face interaction times for participant pairs. Two-sided interactions refers to interactions where both participants are facing each other, while one sided interactions refers to interactions where at least one participant is facing the other, but the other participant may be looking in a different direction.

that there was not enough space for participants 1 and 4 or participants 2 and 3 to rotate their chairs such that they could face each other, as seen in Section 3.6. Taking into account face-to-face interactions in which only one participant faces the other significantly changes the total interaction times for each participant pair and the relative interaction times when comparing pairs. However, it does not significantly change the ordering of which participant pairs interacted most often. In total, four of the participant pairs differ by one rank, and the Spearman's rank correlation coefficient between one-sided and two-sided face-to-face interactions is 0.95, indicating that the interaction time rankings between one-sided and two-sided face-to-face interactions is highly correlated.

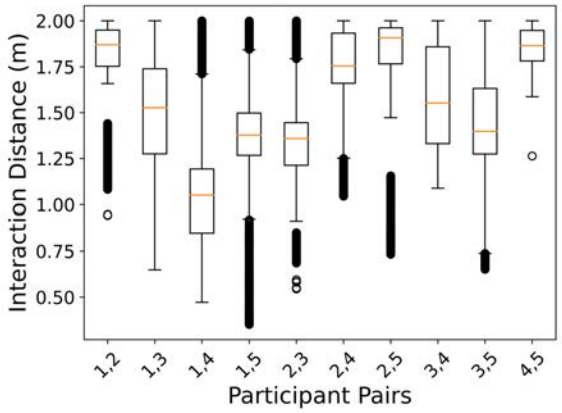
Body Worn Sensors moderately reflect Face-to-Face Interactions

We find that the distribution of body-to-body interaction distances from Opo generally matches face-to-face interaction distances from the Optitrack system, as seen in Figure 3.8a and Figure 3.8b/Figure 3.8c. The Opo interaction distances are generally slightly shorter than the Optitrack distances, which is expected since the Optitrack system measures distances from the centers of participants' heads while Opo measures from the front of their chests. However, we note that in some pairs, such as participants 4 and 5, the median interaction distance and general interaction distance distribution varies significantly. This leads to a moderately strong Spearman's rank correlation coefficient of 0.50 even though visually the body-to-body and face-to-face interaction distances largely match for most participants.

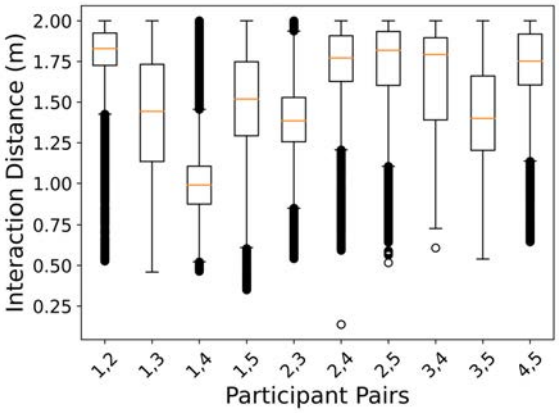
Comparing Figure 3.8d with Figure 3.7a and Figure 3.7b, we see that body-to-body interaction times correlate with face-to-face interaction times. When comparing interaction



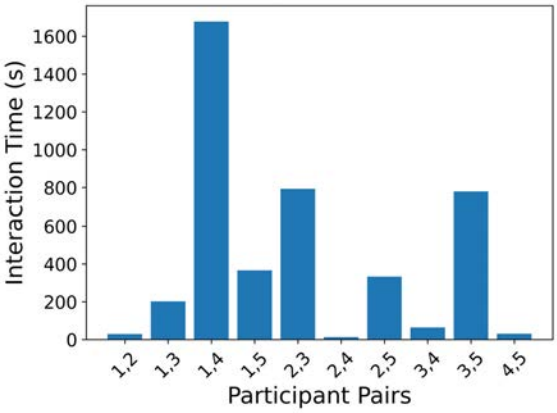
(a) Interaction Distances, Opo



(b) Interaction Distances, Face-to-Face (FTF)



(c) Interaction Distances, One Sided FTF



(d) Interaction Times, Body

Figure 3.8: Comparing body-to-body and face-to-face interaction times and distances. Body-to-body based interaction distance and time is generated from Opo sensors, while face-to-face data is from the Optitrack motion capture system. Figure 3.8b shows interactions in which both participants face each other, while Figure 3.8c show interactions in which at least one participant faces the other. Overall, Opo slightly underestimates face-to-face interaction distance. This is expected because the Optitrack system measures the distance between the center of each participant's head rather than the front of their chests.

time rankings, body-to-body interaction times have a strong rank correlation coefficient with both two-sided face-to-face (0.85) and one-sided face-to-face (0.81) interactions. However, the actual interaction times between body-to-body interactions and face-to-face interactions differs significantly, with high variance between different participant pairs. Thus, we believe it would be difficult to accurately estimate true face-to-face interaction times using body-to-body interaction times.

3.8 Summary

Overall, we believe that our office study shows that the high-resolution, directional, walled off interaction time and distance data from Opo can provide significant value over current proximity based solutions that are often used for disease modeling and tracking. Furthermore, preliminary results from our motion capture study indicate that even in restrictive environments that discourage people from fully turning their bodies to interact with one another, Opo and other chest-worn directional sensors can still serve as a proxy for face-to-face interactions. While fully capturing face-to-face interactions in these restrictive environments likely requires actual head-mounted gear, we believe these results show that overall, Opo and other chest-worn directional interaction sensors are flexible enough to be deployed in a variety of environments.

While understanding face-to-face interaction networks is critical to understanding how diseases such as COVID-19 are spread, they are not the only components of interactions. Moving forward, we explore how we can extend Opo and utilize passive vigilance to track health-behaviors and symptoms. Specifically, we explore sensing hand-washing due to its near universal importance in combating disease-spread, and sensing coughs, one of the most common symptoms of respiratory illnesses.

Chapter 4

Tracking Hand Washing Rates

4.1 Motivation and Background

Hand washing is one of the most important ways to prevent disease spread, and hand hygiene compliance has gone hand-in-hand with contact tracing as two of the most important tools in combating the COVID-19 pandemic [76, 77]. Epidemiologists have put significant effort into understanding both hand washing rates, or how often people was their hands, and hand washing techniques, or how well people follow recommended guidelines. Surveys offer the most convenient way to study hand washing rates and techniques, but their results are highly inaccurate, making secret observers the current standard for acquiring this data [78].

Due to hand washing's role in preventing healthcare associated infections, most secret observer studies have been done in hospital settings. These studies have found that hand washing compliance varies greatly between different institutions, but on average hand washing compliance is only around 38% among health care workers in both developed and developing countries [79]. Hand washing compliance takes into account both hand washing rates and hand washing technique, but isolating hand washing rates produces similarly concerning results. For example, a study of six intensive care units (ICUs) found that physicians and nurses only washed their hands around 65% of the time after contact with fluid or bodily excretions, and had similar hand washing rates before using an invasive device on a patient [80]. Hand washing compliance among different institutions has also been found to be highly varied. For example, among the six ICUs previously mentioned, hand washing compliance ranged from 3% to 100% [80]. Overall, these results have led to policy makers to conclude that periodic, if not constant, monitoring of hand washing in hospitals to be an important tool in combating health care related infections [78, 81]. Outside of hospitals, secret observers have been used to study hand washing behavior in a variety of public restrooms. These studies have found that hand washing rates are trending upwards, show a strong gender bias with women washing their hands far more frequently [82, 83]. Overall, in public restrooms, people tend to wash their hands around 85% of the time, ranging from around 75% to 90% depending on the restroom observed [82, 83]. Observed hand-washing times are almost universally under 15 s,

and only about 65% of people used soap [83].

However, outside of a few studies on public restrooms, we have a limited understanding of hand washing behavior outside of hospitals. Conducting secret observer studies in many contexts is simply impractical. While a secret observer can hang around a few public restrooms used by a variety of people without arousing suspicion, doing the same in employee bathrooms in an office or restaurant is much more difficult. Furthermore, even when secret observer studies are feasible, they are costly, do not cover all hand washing stations, and only provide insights into group behavior. Studies on public restrooms typically only cover a subset of restrooms at a location, and in hospitals studies typically only cover one department or a subset of hand washing stations. Tracking individual hand washing behavior is also impractical for secret observer based studies, since this would require the observers to be able to identify the people they observe. While secret observer studies can give us aggregate group statistics, they cannot offer insights on individual behavior, such as how common super spreaders that never wash their hands are.

These limitations have led both academic and industry stake holders to seek sensor based solutions to provide objective, automated, and individualized hand washing data. Academic work has largely focused on using accelerometers and other sensors in smart watches to detect a hand-washing motion [84–88]. This technique allows these systems to detect not only hand washing events, but also hand washing duration. However, smart watch based solutions cannot be used in situations where wrist worn jewelry is considered a safety hazard, such as food handling and certain healthcare settings, limiting the situations they can be deployed in [89–92]. Furthermore, these systems have no method of detecting if soap was used, which is an important consideration [78, 81, 83]. Commercial hand hygiene monitors generally combine smart badges with wall or soap dispenser mounted sensors [93–99]. These systems work by having the mounted sensors detect when a soap dispenser is used, and then detecting what user is in close proximity to the soap dispenser. The key assumption is that if a user is in close proximity to a soap dispenser when it is used, the user is likely washing their hands.

This architecture can be used to capture soaped hand washing events in a wide variety of situations. Unlike wrist-worn systems, smart badge systems typically do not measure hand washing rates. However, these commercial systems are targeted at increasing hand hygiene compliance in hospitals rather than being a flexible hand hygiene monitor. They are often integrated into larger, more complex systems that track when hospital workers should wash their hands and provide live reminders. This makes them more suited for permanent hospital deployments rather than temporary research deployments. Practically speaking, these systems are not available for general purchase, and operate on subscription models, making them difficult for researchers to use. Finally, current smart badge systems only capture hand washing events, not hand washing duration.

Overall, researchers lack a practical sensor solution to answer even the simplest individual hand washing question: how many times a day does an individual person wash their hands with soap? To help address this design point, we prototype an extension to the Opo sensors which allow it to detect soaped hand washing events and hand washing duration. We adopt the smart badge plus soap dispenser sensor infrastructure found in many commercial systems,

extending Opo to detect soap dispenser usage by using accelerometric passive vigilance. In addition, we take advantage of Opo's ranging accuracy to estimate hand washing duration, making it the first smart badge solution to do so.

4.2 Related Work

Academically, work on sensing hand washing has largely focused on using smart watches to detect hand washing. These systems use accelerometers, gyroscopes, and/or inertial sensors to detect when users are making hand washing motions with their hands [84, 85, 87, 88]. In addition to detecting when users wash their hands, some systems aim to detect how well users are following the WHO guidelines on how to was their hands [88]. These systems have shown mixed results but continuous improvements, with the latest system having a precision and recall of about 0.7 using a lab controlled data set of 18 common physical activities [85]. However, these systems are generally untested in the wild, and it is unclear how often people make motions similar enough to hand washing to generate false positives in these systems. In addition, people wash their hands in different manners, which significantly affects person-to-person accuracy without personalised training data [84, 88]. Li et al conducted a 4 hour test where users were allowed to leave the lab but told to wash their hands in a specified sink during the test, and achieved a 12.5% false positive rate in this test [88]. However, participants in the study were trained and asked to wash their hands according to WHO guidelines, making it difficult to determine how well their system works for monitoring untrained people. Despite the fact that users consistently washed their hands according to WHO guidelines, Li et al still found that person-to-person hand washing motions varied highly, with their system accuracy dropping from 85% to 69% when switching form a user-dependent to a user-independent model. To compensate for false positives one system, Harmony, integrates Bluetooth beacons placed on soap dispensers and only evaluates accelerometric/gyroscopic data for hand-washing when the smart watch is able to receive packets from these beacons [84]. The idea is that if people are near a dispenser are making hand washing motions, it is because they are washing their hands. However, they do not evaluate this assumption, nor do they evaluate how well their Bluetooth proximity sensing ensures that users are near the dispenser. Zhang et al take an entirely different approach, and instead of relying on accelerometric or gyroscopic data use a custom built smart ring that integrates a fluid sensor [86]. In lab settings the ring could consistently detect when water covered the ring, but is untested outside of a lab setting.

Other researchers have used sensors which measure soap dispenser usage as a proxy for hand-washing rates [100, 101]. However, it is unclear what the mapping between soap dispenser usage and hand-washing rates are, and such an approach only gives researchers a general sense of group behaviors. These systems also register each press of a soap dispenser as a hand washing event, which may lead to many false positives if users press the dispenser multiple times. Commercially, Apple recently (Sept 2020) added hand washing detection to the Apple Watch [102]. Apple's hand washing detection relies on both motion and audio

sensing to detect when users wash their hands, although they offer no official accuracy or reliability figures.

Most commercial hand washing systems are not targeted at consumers, but rather at increasing hand hygiene compliance in hospitals. Instead of using wrist worn wearables to detect a hand washing motion, these systems generally combine smart badges with wall and/or soap dispenser mounted infrastructure [93–99]. Most of these systems localize smart badges against mounted infrastructure to determine when a user is in close proximity to a soap dispenser. If a soap dispenser is used, nearby user(s) are assumed to be washing their hands. Notably, proximity is not used to estimate hand washing duration. These hand washing sensor systems are integrated into a more complicated system which attempt to detect when users should wash their hands (e.g, when they enter a room) and/or provides live visual or audio reminders. An exception to this architecture is BioVigil, which relies on users manually tapping their smart badge to indicate that they have washed their hands [93]. In addition to a smart badge, Vitalacy also offers a smart wrist band solution which tracks hand washing duration in addition to hand washing rates, although they do not specify how it does so [98].

Technologically, these systems appear to rely on Bluetooth or WiFi proximity sensing for localization, although none of the vendors we found specify exactly how their systems work [95, 98, 99, 103, 104]. It is unclear how accurate their localization techniques or hand washing rates are, and accuracy rates are not advertised by vendors. Dyson and Madeo interviewed nurses using automated hand hygiene monitors, and found that inaccuracy, both in detecting when a nurse should wash their hands and if hand washing occurred, was a common complaint [105]. It is important to note that the goal of these systems is not to accurately monitor hand hygiene behavior. Rather, these systems aim to increase hand washing rates in hospitals, and evidence suggests that they do accomplish that [105, 106].

4.3 System Design

Our Opo hand-washing sensor system emulates the wearable smart badge + soap dispenser attached sensor architecture found in many commercial systems. This architecture uses two steps to determine when someone washes their hands: detect when a soap dispenser is triggered or used, and then detect what users are in close proximity to the soap dispenser. Users in close proximity to the soap dispenser when it is used are assumed to be washing their hands with soap. In addition to capturing when users wash their hands with soap, our system captures un-soaped hand-washing events and hand-washing duration. To the best of our knowledge, no other smart-badge based system captures hand-washing duration or un-soaped hand-washing events.

We posit that this is because current systems use BLE based proximity sensing or other similarly low-resolution, omni-directional ranging technologies to determine how close smart-badge wearers are to a soap-dispenser attached sensor. These technologies are typically only accurate to within 2-3 m [34]. Because of this, they cannot accurately detect when

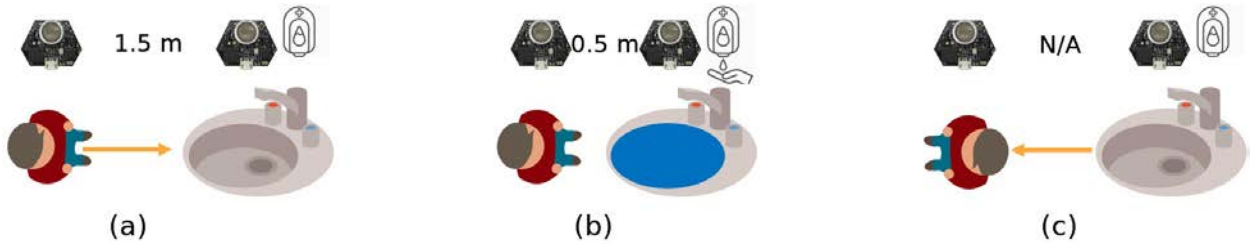


Figure 4.1: System overview. The person shown is wearing an Opo sensor, and an Opo sensor is mounted to the soap dispenser. The person faces as the soap dispenser as they walk closer to wash their hands (a). When they get to the sink to wash their hands, the Opo sensors detect an extremely close distance between the person and the sink (b). In addition, when the person uses the dispenser, it triggers an accelerometer on the dispenser mounted Opo. Our system determines that this person is washing their hands due to their close proximity to the soap dispenser when it is used. When the person is done washing their hands, they walk away (c), resulting in either longer distance measurements between the person and dispenser Opos, or a lack of distance measurements between them. This lets the system know that the person is done washing their hands, and the total hand washing duration is estimated as the time between the soap dispenser trigger and the last close range measurement between the two Opos.

people are close to the dispenser just by using their proximity sensing, and rely on both the soap dispenser being used and proximity sensing to ascertain that a smart badge wearer is near a sink. Similarly, the low-resolution of their proximity sensing means that they cannot accurately detect when a user finishes washing their hands and leaves the sink. In contrast, Opo provides high-resolution, directional distance measurements between a soap dispenser attached Opo and a Opo wearing user. Thus, we are able to determine when an Opo wearer approaches and leaves a soap dispenser, and use this to estimate hand-washing duration, as shown in Figure 4.1. In addition, because we separate out soap-dispenser usage from detecting when a person is near the dispenser, we can detect both un-soaped and soaped hand-washing events. Our key assumption is that when someone is near a soap dispenser, the vast majority of the time it is to wash their hands.

Attaching Opo sensors to both soap dispensers and users allows us to accurately determine when users are near the dispenser, but do not tell us when if and when the dispenser is used. To detect when a soap dispenser is triggered, we add an accelerometer capable of triggering an interrupt when an accelerometric threshold is reached. We observe that when a soap dispenser is idle, it is extremely still. However, when a soap dispenser is used, the case will either flex or vibrate depending on if the dispenser is an automatic or manual dispenser, causing a corresponding spike in the acceleration an accelerometer attached to a soap dispenser experiences. When the accelerometer is triggered, the soap dispenser attached Opo records the time the dispenser was used. In testing on automated and manual wall mounted soap dispensers, we found that it was trivial to detect soap dispenser usage using

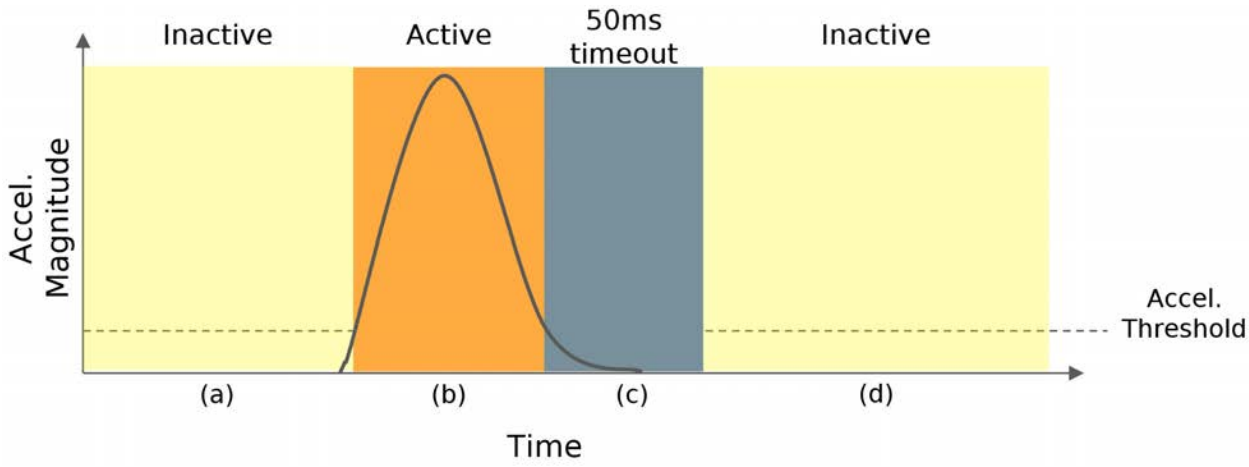


Figure 4.2: Detecting soap dispenser usage using an accelerometric trigger. Opo's accelerometer has two phases, inactive and active. When the soap dispenser is not in use, the accelerometer reads a negligible acceleration (a), and is inactive. When the accelerometer is used, the accelerometer experiences a spike in acceleration (b), causing the acceleration to exceed the programmed threshold (0.032 g). This causes the accelerometer to switch to the active state and trigger an interrupt to the Opo's mcu, letting the mcu know that the dispenser was used. After the acceleration falls below the threshold for 50 ms (c), the accelerometer assumes the person is done using the dispenser, and reverts back to the inactive state (d).

a simple accelerometric threshold of 0.032 g, as shown in Figure 4.2. In testing on local wall mounted manual and automatic dispensers, we found that such a threshold was able to detect 100% of soap dispenser activations, even when we pressed the manual dispensers in an unreasonably slow fashion. Moreover, washing our hands, walking near, or even jumping up and down near the soap dispenser did not trigger any false positives.

While our system is deployed, the Opos simply store distance measurements between users and soap dispenser and timestamps of when soap dispensers are used. In post processing, we analyze this data to determine when users washed their hands, hand-washing durations, and whether or not soap was used. To do so, we first set two parameters, `dispenser distance` and `time buffer`. `dispenser distance` is the maximum distance we expect a person to be from an Opo attached to a soap dispenser while washing their hands. This must set per deployment or possibly per soap dispenser depending on the deployment. `Time buffer` is the maximum time we expect between Opo ranging measurements. In theory this is simply how often Opo sensors attempt a ranging event, or around 3 s. However, in practice we found that Opos may miss a few events for a variety of reasons, and believe 7 s is a reasonable `time buffer`.

We separate out Opo distance measurements into user and soap dispenser pairs and filter out distance measurements that are greater than `dispenser distance`. We then split up

the distance measurements into hand-washing events by looking for sequences of distance measurements between the user-dispenser pair where each distance measurement is at most `time buffer` seconds later than the previous distance measurement in the sequence. This gives us both hand-washing times and durations between that user and soap dispenser. To determine if a hand-washing event is soaped or un-soaped, we look look to see if the soap dispenser is used within `time buffer` seconds of the first distance measurement in the hand-washing event. Finally, we filter out hand-washing events such that each user-dispenser pair only has one hand-washing event per 2 minutes, prioritizing soap use and duration, in that order, when selecting what hand-washing event to keep when multiple events are present. This step filters out hand-washing events from being double counted when there are multiple dispensers at a sink.

4.4 Evaluation

To evaluate our system, we first tested our system on wall mounted soap dispenser near our lab. We found that it was trivial to get the system working perfectly. However, it was unclear if the accelerometer based soap dispenser trigger was over fit to our local soap dispensers. To get a better idea of if our system was a feasible, we conducted a deployment at a Midwest office site. We first check if our time buffer of 7 s is reasonable and then check to see if our overall hand-washing data is reasonable. Ideally, we would have ground truth through a simultaneous secret observer study. However, for logistical reasons, this proved impractical in our office setting. Instead, we compare our data against data from a prior secret observer study from Borchgrevink et al in public restrooms [83]. We select this study because it is the only one we found that contains information on hand-washing duration and soap use outside of a hospital setting. Our general hypothesis is that if our system worked correctly and 7 s a reasonable time buffer, our data would roughly match the data found in this study.

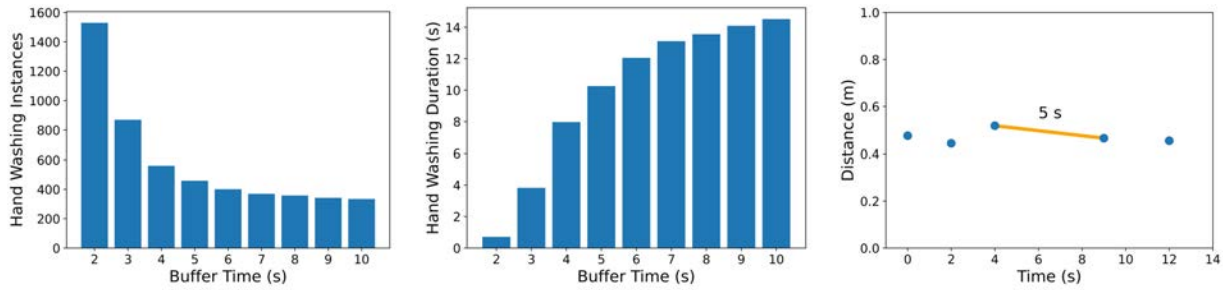
Study Design

To test our system, we incorporated our hand-washing detection system into the two office deployments described in Section 3.3. As previously discussed, office site B had ultrasonic lighting control systems that interfered with Opo's operation, so we focus our analysis on office site A. At office site A, we deployed Opo sensors on 23 people at an office site for a period of five weekdays. In addition, we deployed Opo sensors on three Purell sensors and 16 soap dispensers. 14 of the soap dispensers were located in bathrooms, while two were located in common break areas. 22 out of the 23 people had at least one hand-washing event, 21 of which had at least one hand-washing event with soap. Manually checking the data and periodic checks during the deployment did not indicate any deployment problems.

Because bathrooms contained multiple soap dispensers, and thus multiple soap soap dispenser Opos, participants could be in close proximity to multiple soap dispenser Opos while washing their hands. To account for multi-counted hand-washing events when comparing

our data to prior work, we time filtered hand-washing events. Each participant can only have one hand-washing event per two minutes, with a preference for soaped and longer hand-washing instances when selecting which hand-washing event to count. After time-cleaning our data, we had 199 hand-washing events, or about 1.8 hang washing events per person per day. We found no problems with our soap use detection system in office A, but note that in office B one bathroom shared a wall with a utility room. The wall and the soap dispensers mounted to this wall both constantly vibrated, resulting in our accelerometer based soap use system constantly triggering false positives in this bathroom.

Time Buffer Exploration



(a) (b) (c)

Figure 4.3: Comparing the effect of different time buffers on the number of sensed hand washing events (**a**) and mean sensed hand washing duration (**b**). (**C**) shows an example hand washing trace. The blue dots are Opo ranging measurements between the person's and the dispenser's Opos, and are the distance measurements recorded during a 12 s hand washing event. Because our method of cleaning the data homogenizes the number of events regardless of the selected time buffer, (**a**) is generated from data that has not been time-filtered to show the effect of changing the time buffer.

While in theory Opo measurements should occur between 1-3 seconds apart, in practice we have found Opos can occasionally miss packets due to wearability problems, environmental noise, or multipath reflections of ultrasonic signals. Based on our previous experiences with Opo, we expected that a time buffer of around 7 seconds would be suitable to account for this.

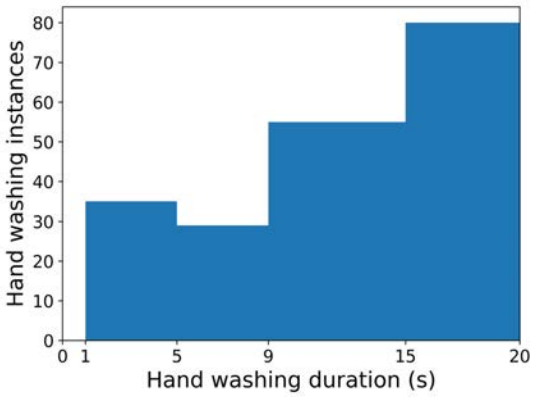
To figure out what an actually suitable time buffer would be, we compare the effect of time buffer values on the number of detected hand washing instances and mean hand washing duration (Section 4.4). When the time buffer is set too low, the number of sensed hand washing events is artificially high and the sensed hand washing durations are artificially low. To illustrate why, we examine an example trace from the sensors during a hand washing event in Figure 4.3c. Here, we see that while most of the Opo ranging measurements are

separated by two or three seconds, the third and fourth ranging measurements are separated by five seconds. In this case, a time buffer of less than five seconds would separate these five measurements into two hand washing events of four and three seconds, rather than a single twelve second hand washing event. Intuitively, these sensor readings likely represent one hand-washing events, rather than two hand washing events that occur within the span of 12 seconds. Thus, when the time buffer is too low, we see many short hand washing events that occur in a short amount of time, rather than a single hand washing event, which results in an artificially high number of hand washing events and an artificially low average hand washing duration. As we increase the time buffer, the system starts coalescing short "separate" hand washing instances that occur closely in time to singular longer hand washing events. This results in the number of sensed hand washing instances falling, while the interaction durations rise as short "separate" hand washing instances are coalesced into single longer hand washing instances.

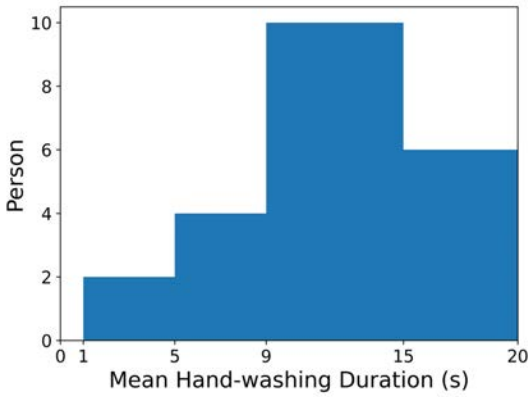
Manually examining the data, it does appear that a 7 s time buffer is sufficient to largely group the data correctly. Furthermore, as we can see in Figure 4.3a, at a 7 s time buffer, the number of detected hand-washing events and that average hand-washing duration has largely smoothed out, indicating that a 7 s time buffer does a good job of grouping together Opo interactions located closely in time into one hand-washing event. Only 35 instances (9%) and a 1.4 s difference in mean hand-washing duration (10%) separates using a 7 s and a 10 s time buffer. Increasing the time buffer from 7 s does not significantly change the distribution of our data or our overall conclusions. Thus, we believe a 7 s time buffer is a reasonable selection.

Comparing to Secret Observer Studies

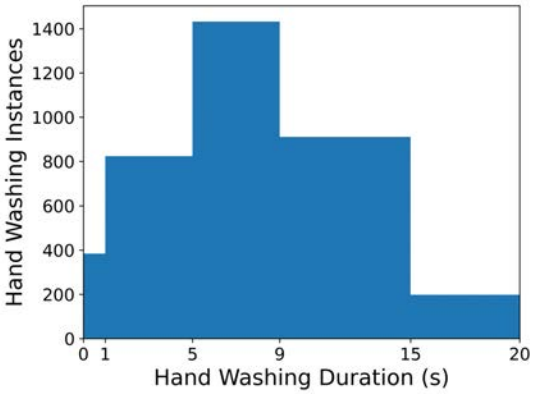
To check if our data is valid, we conduct a sanity check by comparing our data to Borchgrevink et al [83]. Borchgrevink et al conducted their secret observer study in public bathrooms located across a college campus and college town. The researchers do not specify what college town their study is conducted in. However, the authors are all part of Michigan State University in Lansing, Michigan. In total, they observed 3365 hand-washing instances and an additional 384 instances of people using the bathroom and not washing their hands. Due to the nature of the study, keeping track of participants is impossible. Instead, Borchgrevink et al treat each hand-washing instance as belonging to a separate participant, assuming that their study design and size prevents multi-counting from significantly influencing their results. Our system is able to track participants, and our study design produces different results depending on if we look at hand-washing instances or participants. Thus, we show both hand-washing data by instance and participant when possible. In addition, we note that our study demographics are different from those of Borchgrevink et al. However, because the full data set in their study is not available, we largely cannot demographically adjust their results. In particular, we believe that our cohort should have longer hand-washing durations than Borchgrevink et al's cohort, and discuss this in more depth later.



(a) Instances



(b) Participants

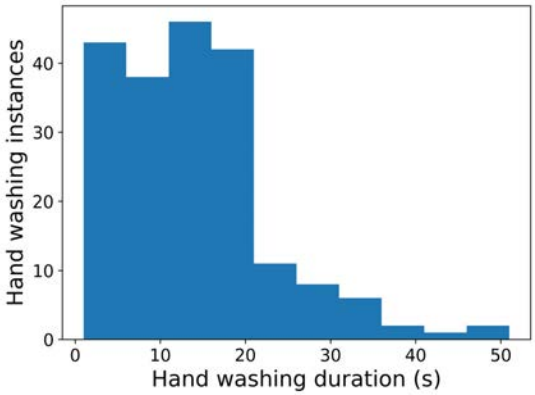


(c) Prior Work [83]

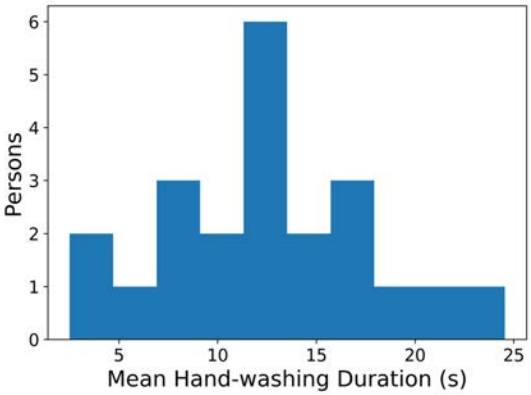
Figure 4.4: Histograms of Opo hand-washing data and prior work. **a** is a histogram of all hand-washing instances captured by Opo, while **b** shows a histogram of mean hand-washing data per participant in our study. **c** shows the distribution of prior work. The histogram bins are the ones used in Borchgrevink et al's paper [83]. The 15-20 s bin contains all hand-washing durations above 15 s. The 0 s hand-washing duration bin represents people who did not wash their hands.

We first compare the distribution of our hand-washing data and prior work. Borchgrevink et al only provide non-uniformly binned data, as shown in Figure 4.4c. When we bin our hand-washing instances in the same way, a shown in Figure 4.4a, our distribution looks significantly different from the prior work distribution (Figure 4.4c). When we look at the distribution of the participants' mean hand-washing duration (Figure 4.4b), we see that our distribution shape much more closely matches the prior work (Figure 4.4c), albeit with a higher mean. Quantitatively, 69% of hand-washing instances in Borchgrevink et al's study are between 5-14 s long. While only 42% of our hand-washing instances are between 5-14 s long, 59% of participants have a mean hand-washing duration between 5 and 14 s. Overall, the

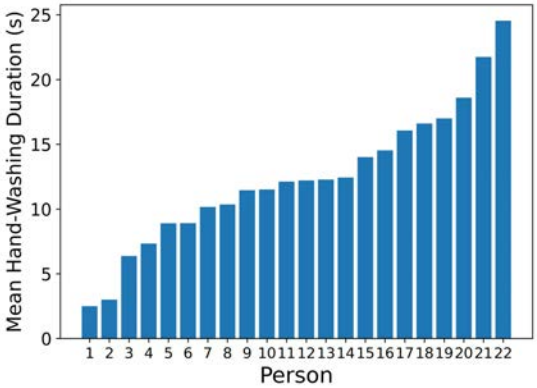
fact that participant-normalized data matches prior work so well indicates that our system generally does a good job measuring hand-washing data. In addition, the differences between hand-washing instance and participant data shows the benefits of using a wearable sensor system instead of a secret observer based study, where participants and events are considered the same.



(a) Histogram, Hand Washing Instances



(b) Histogram, Participants

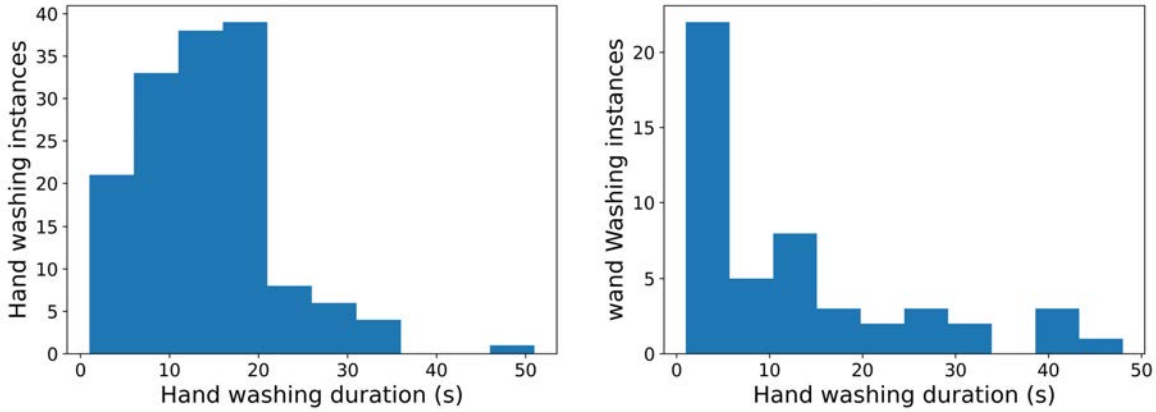


(c) Participant Breakdown

Figure 4.5: Data distribution and breakdown for our study. The bins used in Borchgrevink et al's study present a misleading representation of our data, so we present more accurate histograms in **a** and **b**. IN addition, **c** shows a breakdown of all our participants. As we can see, participants 21 and 22 have a significantly higher mean hand-washing time than other participants. Based on the length of some of their hand-washing instances (40+ s), we speculate that they were doing things at the sink other than washing their hands.

The histogram bins used by Borchgrevink et al are a poor fit for our data, and present a misleading view of our data. Figure 4.5 shows histograms of our data with uniform bins that fit our data much better. As Figure 4.5a, our hand-washing instance data is actually a right-tailed distribution. Our participant hand-washing data distribution does not change as

drastically, although Figure 4.5b is slightly more normal than Figure 4.4b. Borchgrevink et al's analysis implicitly assumes their data is normally distributed, so we consider the normal shape of our per person data further proof that our system works.



(a) Soap Used (b) No Soap Used

Figure 4.6: Opo hand washing durations, with and without soap. Soaped vs un-soaped hand-washing rates lines up very well with past secret observer studies. Hand-washing durations largely match our intuition that people who do not use soap generally just quickly wet their hands.

Finally, Section 4.4 shows the distributions of our soaped vs un-soaped hand-washing instances. Although Borchgrevink et al provide no duration statistics on soaped vs un-soaped hand-washing, intuitively we expect soaped hand-washing instances to be longer and un-soaped hand-washing events to largely be a brief hand-wetting. Section 4.4 largely validates this intuition, although some of the hand-washing instances are un-realistically long for both soaped and un-soaped events. We speculate on that below.

Means, Demographics, and Soap Use

As Table 4.1 shows, the proportion of hand-washing instances that used soap and that did not use soap is remarkably close between our data set and prior work, with our data only showing a slightly higher rate of using soap when washing hands. However, our mean hand-washing durations appear to be significantly higher than expected whether we are looking at hand-washing instances or participant data. We believe this is due to a few reasons.

First, the population mean hand-washing durations for prior work are not demographically adjusted. While Borchgrevink et al provide a population mean that excludes people who do not wash their hands, their analysis of demographic and other factors includes people who do not wash their hands as having a hand-washing duration of zero. This makes it impossible to demographically adjust their overall mean hand-washing duration. Their study indicates that factors such as age, sink condition, and automatic vs manual faucets can all influence

Data Set	Mean Duration Per Instance	Mean Duration Per Person	Mean Duration Per Person, Male	Mean Duration Per Person, Female	% No soap Per Instance	% With Soap Per Instance
Prior Work [83]	7.52 s[a]	7.52 s[a]	6.27 s[b]	7.07 s[b]	26.7	73.3
Opo	13.20 s	12.39 s	12.46 s	12.31 s	24.6	75.4
Opo, Filtered[c]	11.80 s	11.32 s	10.32 s	12.31 s	23.3	76.7

[a] Means exclude people who did not wash their hands.
[b] Means include people who did not wash their hands as having a hand-washing duration of 0 s.
[c] Excludes two participants with significantly higher hand-washing durations than other participants.

Table 4.1: Comparison of mean hand-washing duration and soap use rates between our data and prior work [83]. Overall, we find that our soap use rates are remarkably consistent with prior work. However, our mean-duration is higher, even once we remove two outlier participants with suspiciously high mean hand-washing durations.

hand-washing duration. Specifically, they found that being older than college age, having a clean sink, and having automatic faucets all increase mean hand-washing duration, and our study has an older cohort, universally clean sinks, and automatic faucets. Based on this, we speculate that our expected mean-hand washing duration is longer than the 7.52 s listed.

Second, our data is heavily influenced by two participants with mean hand-washing durations longer than 20 s, as shown in Figure 4.5c. Our system recorded 23 hand-washing instances between these two people, resulting in a significant impact on our mean hand-washing duration. Excluding those two participants drops our mean hand-washing instance duration to 11.80 s and our participants' mean hand-washing duration to 11.32 s.

Finally, we speculate that our system is capturing events outside of hand-washing. For example, participants 21 and 22 in Figure 4.5c have multiple hand-washing instances longer than 40 s. We speculate that it is highly unlikely that these events are actually everyday hand-washing events, and we may be capturing other events such as teeth brushing or washing off a stain. In addition, even when people are washing their hands, they may also spend a few seconds checking the mirror, flicking water off their hands, etc. These events would increase our systems' sensed hand-washing duration, but would not increase a secret observer's measured hand-washing duration.

4.5 Summary

Overall, we believe the distribution of our hand-washing participant data, our soap and no-soap hand-washing rates, and our gender differences are promising evidence that our system accurately captures hand-washing trends. Our system and other smart-badge based systems use proximity and soap-dispenser use as proxies for hand-washing. This study indicates that the majority of the time, this is likely true, but hand-washing durations may be over-estimated by people spending a few seconds at the sink before or after washing their hands. In addition, we likely captured a few events in which participants were using the sink for purposes other than hand-washing.

We believe that these results also show that the Opo ranging primitive can be useful beyond

face-to-face interactions. Extending Opo with interrupt generating accelerometers or IMUs may provide a general method of allowing Opo to detect human-object interactions, which can provide important data in a variety of applications. Of course, short-range interactions between people and people, and people and things, are not the only epidemiologically important events that can be tracked. In the next chapter, we explore the potential for practically detecting coughs, both a symptom and transmission vector of many respiratory illnesses.

Chapter 5

Privacy Preserving Cough Sensing

5.1 Motivation and Background

Coughing is a common symptom that results in significant health care costs and quality of life reductions for many people [107]. In the USA, coughs are the fourth most commonly cited reason for physician office visits, behind only general check ups, post-op check ups, and other progress visits [108]. In addition, coughs are a symptom of over 100 pathological conditions, often serving as an early warning sign and a potential infectious vector [109]. Most recently, coughing is one of the most commonly cited symptoms of the global COVID-19 pandemic.

Because of its effects on patients, its significance as an early warning sign of many diseases, and its potential as an infection vector, coughs are an are of interest in many public health studies and in real world health policy. Typically, cough data is collected simply by having people self-report the subjective onset of cough symptoms [110–113], although more detailed cough surveys also exist [112, 114]. For example, in a six year study on the effectiveness of influenza vaccines, researchers asked vaccinated participants to self-report the onset of symptoms including coughs, after which they were tested for influenza [113]. More recently, coughing is one of the symptoms used to prioritize which people should get tested for COVID-19 in the absence of sufficient tests for everyone.

Although self-reporting coughs is a convenient way of collecting data, self-reporting is well know to result in coarse-grained data of questionable accuracy, affecting researchers' abilities to rely on cough symptoms [115, 116]. For studies that rely on self-reporting cough onset, it is difficult to know if a lack of active communication from participants is a sign of being asymptomatic or non-compliance. Furthermore, if participants report in at a later date, it is unclear how reliable their memory of exact symptom onset is.

These problems have driven efforts to automated wearable cough counting systems for over a decade [117]. These systems prioritize providing accurate cough counts, or the number of times a user coughs in a day, and have shown promising results in small scale tests [118–120]. However, to acquire the needed audio data, participants are required to constantly record their audio environment, with the audio being analyzed in post processing for coughs. This

results in both usability and privacy complications. Users are required to wear a sound recorder large enough to record at least an entire day's worth of audio, wired to a lapel microphone for optimal cough audio quality. More worryingly, using these systems presents a massive privacy intrusion, and legal obstacles in regions with dual-party consent laws [121].

The privacy implications of most cough counting systems have led researchers to explore utilizing smartphones to either locally process an audio stream for coughs [122], or locally extract audio features that allow coughs but not speech to be reconstructed during post processing [123]. However, locally processing the a smartphone's audio stream for coughs takes 25% of the smartphone's battery life over the course of 16 hrs [122], and locally extracting smarter audio features was never implemented on a smartphone [123]. Modern smartphones also do not allow apps to continuously monitor the microphone in the background, making these techniques currently infeasible to implement.

In addition, while highly accurate cough counts have interesting uses, such as monitoring tuberculosis recovery [124, 125], they are not required to provide value. More commonly, studies are simply interested in knowing when participants begin or stop coughing, and possibly if the coughing is getting worse or better [110–113]. Sick patients often cough an order of magnitude or more than healthy ones, meaning that even a very approximate cough count can provide an objective and automated method of acquiring the data these studies seek [126, 127].

These factors point to the need for a specialized cough recorder that prioritizes privacy as an alternative to using an always-on voice recorder. Motivated by this design point, we present CoughNote, a 1.3 x 1.3 inch wearable cough recorder that seeks to address the need to continuously record audio. CoughNote utilizes a novel hardware trigger which activates for the start of a potential cough, after which the system records a 1 s audio clip, which can then be analyzed in post processing. Crucially, the trigger is designed not to be triggered by speech or other vocal sounds, heavily minimizing the chances of catching any sensitive vocal audio. To evaluate CoughNote, we conduct a small pilot test with high-fidelity ground truth, finding that CoughNote captures 65% of coughs, and generates an average of 8.3 false positives per hour, or 2.2 min of wasted audio per day. We further provide evidence that show that CoughNote generalizes to coughs from other people and causes by evaluating CoughNote on four patients infected with influenza, finding that CoughNote is able to capture a variety of cough sounds, and correlates strongly (r=0.91) with self-reported cough count estimates. CoughNote accomplishes all of this while preserving user privacy, and being smaller, lighter, and almost three times as long lived as a typical voice recorder.

5.2 Related Work

Coughs are commonly tracked by having participants self-report the onset of coughing to researchers through participant initiated communication, periodic surveys, or other forms of self reporting [110–113]. Although normally only coarse-grained information about coughs is required, specialized surveys are used to gather more detailed cough data [114]. It is well

known that surveys result in subjective data of questionable accuracy [115]. Even when only coarse grained information is required, the subjectivity of self-reported methodologies can prove problematic. It is unclear if users not actively self-reporting coughs are asymptomatic, or are simply not reporting their symptoms. Similarly, periodic surveys rely on human memory to determine the onset of coughs.

Due to the these issues, numerous attempts have been made to design ambulatory, automated, and objective cough coughing systems. These systems prioritize providing accurate cough counts, or the number of times a user coughs in a day. provide cough frequency as an objective measure of cough. Commonly cited reasons for developing cough counters include potential early detection of various respiratory diseases and for monitoring treatment efficacy [114, 116, 118, 123]. Generally speaking, these systems aim to provide accurate cough frequencies, to last at least 24 hours per battery charge, and cause minimal inconvenience to users [116].

In this section, we show that while these systems have shown promising results in small scale studies, collecting the audio needed to gather coughs in the wild remains an important unsolved problem. In addition, we evaluate many of the motivations of previous cough counting sensors, and find that most of the value of cough counting can likely be obtained with approximate cough counts.

Automated Cough Counters are Highly Intrusive

Broadly speaking, cough counters can be divided into contact and non-contact systems. Contact systems adhere sensors such as accelerometers directly to a users' skin, are significantly bulkier than non-contact systems, and are no better at detecting coughs [114, 128]. As a result, most work in cough counting focuses on non-contact monitoring.

Non-contact cough counters instead analyze audio traces for coughs. Typically, users carry a portable voice recorder and attach a lapel pin microphone to their shirt, continuously record their audio environment, and researchers then label coughs in a portion of these recordings, which are then used to train models which can detect and count coughs [118, 120, 129, 130]. These systems have worked well in limited testing, with Swarnkar et. al achieving a sensitivity of 93.44% and specificity of 94.52% in controlled tests. However, continuously recording audio presents a major privacy concern that lowers the practicality of audio based cough counting, and may present legal obstacles in areas with dual party consent laws [121].

To address these concerns, researchers have explored methods that would not require all audio to be recorded and sent to the cloud. Larson et al presents a system which relies on a smartphone constantly recording audio, and extracting features which allow coughs but not speech to be reconstructed in the cloud [123]. SymDetector moves all processing to the smartphone, continuously monitoring a phone's microphone and locally processing the data to determine cough counts [122]. However, SymDetector drains roughly 25% of a smartphone's battery life over 10 hours, and 77% of the energy usage is simply from continuously taking audio symptoms. Although Larson et al do not implement their system on a smartphone, Sun et al implement it to compare SymDetector against, and find that Larson et al's system

would drain at 50% of their test smartphone's battery over 10 hours. Furthermore, modern smartphones do not allow third party apps to continuously monitor the phone's microphone, meaning these techniques are currently impractical to implement.

Thus, while the accuracy of cough counters has steadily improved, collecting the needed audio has remained a highly burdensome task. This limits the practicality of using these systems to get even approximate cough counts, which would be enough to provide value for current studies which rely on active self-reporting to determine the onset, disappearance, and general trend of coughs [110–113].

Cough Counting Motivations

Automatic cough counters present similar motivations: coughs are are an infection vector, can significantly affect quality of life, and are a symptom of many diseases. Therefore, accurate cough counts will allow us to better study, monitor, and combat disease severity, infections, and coughing's impact on a person's quality of life [114, 116, 118, 123]. Because the link between cough frequency and quality of life, disease severity, and infectiousness feel so intuitive, it becomes easy to justify focusing on prioritizing cough count accuracy over privacy. Indeed,there is evidence that accurate cough frequencies can be greatly beneficial in some important applications, such as monitoring TB recovery [124, 125]. Anecdotal evidence of cough frequency reductions in a cystic fibrosis clinical trial [131] has also led to speculation that cough frequency can be used as an endpoint in certain clinical trials [123]. In addition, there are many areas where cough frequency has not yet been explored, such as its relationship to air pollution [132].

However, the current corpus of medical and epidemiological literature on coughs shows little evidence that higher cough counts significantly impact quality of life, infectiousness, or disease severity in most cases. Due to the unintuitive nature of these conclusions, as part of our background research, we consulted with two pulmonoglists and an epidemiologist specializing in influenza vaccination to better understanding of why cough counts do not affect these matters. We have included summaries of what we learned from them as supplemental anecdotal evidence where appropriate. We write this section not to call out previous work, but to explain our justification for prioritizing privacy over accuracy, and to emphasize that the primary value cough counters can provide is to support current studies which would benefit from an objective method of determining cough onset, disappearance, and general trends, which does not require particularly accurate cough counts.

Cough Frequency vs Quality of Life

Numerous studies have shown that cough frequency is moderately correlated ($0.3 <= r <= 0.6$) at best with the impact of coughs on a patient's quality of life [114, 133]. While chronic coughs can clearly have a severe impact on quality of life, studies have found that there is significantly more to cough's impact on life than cough frequency [134].

One pulmonologist we consulted explained that in their experience, people do not notice how often they cough. Instead, they notice when coughs disrupt events in their life. So for example, someone that happens to cough once while giving a speech will feel a much bigger effect on coughs on their quality of life than someone who coughs ten times while relaxing at home. Furthermore, this can produce a cyclic effect where patients become stressed that they may cough during an event, increasing the likelihood that they cough, which then triggers further stress. In this way, people can consistently cough only a few times a day, but during situations where a cough is highly disruptive. The pulmonologist also pointed out that even if exact cough counts were available, they would still primarily more attention to how much the patient felt coughs impacted them, since that is a much more useful indicator of how much cough impacts a patient's quality of life.

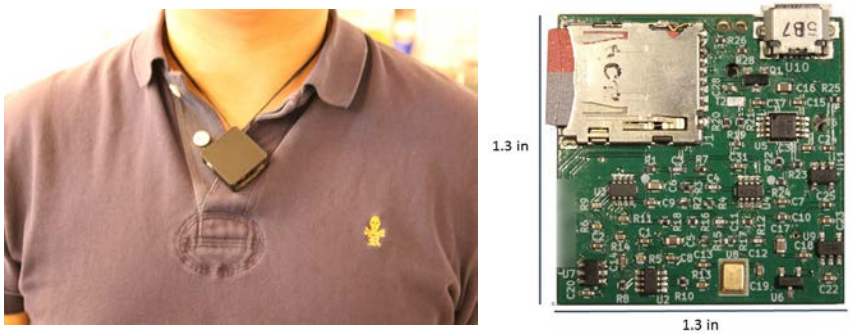
Cough Frequency vs Disease Severity/Diagnosis

Although coughs are a symptom of numerous diseases, we found little evidence that cough frequency is a valid indicator of disease severity or a diagnostic marker. In contrast, cough frequency is not a valid indicator of COPD or asthma severity [135–137]. We even found that one paper cited in a cough monitor design paper as showing a potential link between asthma and cough frequency instead stated "studies have largely shown no relationship between cough frequency and cough receptor sensitivity with various asthma severity indices" [123, 136]. Similarly, while we know that patients with chronic cough or some form of respiratory disease cough more than healthy patients [126, 137, 138], we found no work demonstrating that specific cough frequencies are diagnostically useful. Instead, we found that different causes of cough share significant overlap in cough frequencies [138]. Both pulmonologists we consulted confirmed this, and explained that there is much more to how the body reacts to respiratory diseases than coughs, and the direct effects of a disease on lung function is far more significant than whether or not a cough is triggered.

Cough Frequency vs Disease Infectiousness

Lab tests have found that aerosols from coughs can theoretically transmit influenza [139]. Furthermore, recent studies have shown that aerosols may be a significant influenza transmission vector [140]. However, breathing produces far more aerosols over the coarse of a day [141]. Breath aerosols lack the propulsion of cough aerosols, but can be spread through indoor air ventilation systems [142]. In addition, cough etiquette can significantly impede aerosol spread [143]. Studies directly addressing the link between cough frequency and infectiousness are limited, but they have shown that the relationship between cough frequency and tuberculosis infection rates is non-existent or weak [144, 145].

Although studies in this area have been limited, the epidemiologist we consulted was skeptical of there being a strong link between cough counts and infectiousness. As they explained, not only are there a huge multitude of factors influencing infectiousness, but people may also choose to actively avoid people who are coughing more often, or stay home



(a) CoughNote in case being worn (b) CoughNote pcb

Figure 5.1: 5.1a shows CoughNote in its case being worn as a necklace. The case measures 1.5 x 1.6 x 0.6 in, and the case, battery, and pcb weigh a combined 15.9 g. Figure 5.1b shows our prototype implementation of CoughNote, which measures 1.3 x 1.3 in. The pcb weighs 6.7 g.

themselves if they are coughing heavily. Furthermore, even if there was a link between cough frequencies and infectiousness, it would be incredibly unlikely to affect guidelines on when to self-isolate or proper cough etiquette. Instead, the epidemiologist proposed that we focus on supporting how coughs are currently used in studies, which is as a marker of symptom onset and disappearance, and occasionally symptom trends. This simply requires knowing if a person is coughing significantly more than normal, and being able to measure general trends in cough counts.

Summary

While cough count accuracy can be greatly beneficial to certain applications, we find that in the general case, approximate cough counts that correlate well enough to determine general increases and decreases in cough counts are enough to be greatly beneficial to the way coughs are currently used in many studies. Based on our review of related work and our consultations with medical and health care professionals, we propose prioritizing the user first. What is needed is not a better method of extracting cough counts from audio data. Rather, what is needed is better methods of recording cough audio without massively impacting peoples' privacy. To address this design point, we present CoughNote.

5.3 Design

To achieve our goal of recording coughs without recording sensitive voiced audio, we design CoughNote, a wearable, privacy preserving cough recorder.

Instead of continuously recording audio, CoughNote utilizes a hardware cough trigger which allows our system to continuously analyze an analog audio stream for potential coughs.

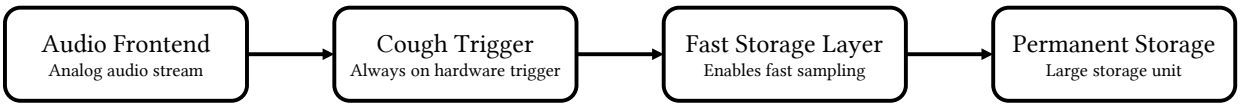


Figure 5.2: System Diagram. The overall architecture of CoughNote is simple. Using a COTS microphone, we stream audio to an always-on low-power (14 uA) hardware trigger. The hardware trigger activates near the beginning of coughs or similar sounds, but crucially does not activate for speech. Once activated, the mcu wakes up, a simple circuit (fast storage layer) enables the mcu to sample and save a 1 s audio clip at faster speeds than it can support with an adc alone to a small temporary external memory bank. Once done recording, the mcu fragments the audio clip based on the mcu's available RAM and transfers the fragments to the permanent storage layer, which in our current implementation is a microSD card for convenience.

Upon detecting the beginning of a potential cough, a 1 s audio snippet is recorded for future analysis. Critically, the trigger is designed in such a way that it is highly unlikely to capture sensitive audio such as talking. As a trade-off of prioritizing privacy, we cannot ensure that all coughs are recorded. In this section, we describe the design of our cough trigger, explain our reasoning for our cough trigger design, and overview the system as a whole.

System Design

CoughNote's system consists of four parts: an audio front end, a hardware based cough trigger, a fast storage layer to enable high audio sampling rates using standard low-cost , and finally a large, permanent storage layer, allowing for thousands of samples to be stored for post processing. These four layers are coordinated by the system microcontroller. The system microcontroller, fast storage layer, and permanent storage layer are normally kept asleep or off, only turning on when the cough trigger activates.

The audio front end outputs an analog signal to the cough trigger and to the fast sampling layer, and consists of a MEMs microphone, whose output goes into two separate op-amp buffers. One op-amp buffer feeds into the cough-trigger, while the other op-amp buffer, feeds into the adc used to record audio snippets. This way, any feedback effects from the cough trigger are isolated, and do not affect the recorded sound snippets. Because our cough trigger utilizes high frequency sound, the microphone used must be an ultrasonic microphone with a positive frequency response up to at least 20 kHz.

The cough trigger layer consists of a hardware trigger that is always on, and when activated alerts the system microcontroller about the start of a potential cough. The trigger consists of a multiple-feedback band-pass op-amp filter theoretically centered around 20 kHz with a Q factor of 20, which feeds into a comparator. Empirically, the filter is centered around 18.6 kHz. The trigger can easily be activated by coughs, but crucially will not be activated by voices, even if the user is screaming at a high pitch. The design of our cough trigger is further discussed in Section 5.3.

When the comparator is triggered, it wakes up the system microcontroller, which wakes up the fast storage layer to capture a 1 s snippet of audio, a duration we selected based on prior work on cough durations [146–148]. The fast storage layer enables a standard low-power microcontroller to easily capture a 1 s clip of 16-bit audio sampled at 40 kHz, which is in line with our requirements and prior work [122, 123]. The fast storage layer is further discussed in Section 5.4.

Finally, once the audio clip is done being recorded to the fast-storage layer, the microcontroller then transfers the audio clip in a raw binary format to a larger, permanent storage layer, such as a miroSD card or a large NAND flash chip. After the data is collected or transferred to the researcher, the raw binary format is then converted to standard audio files. In our current implementation, CoughNote is rate limited to recording at most once every 5 s.

Cough Trigger Design

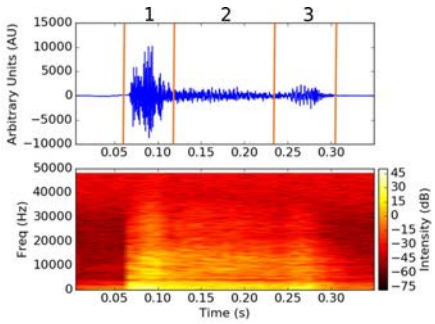
The key component of CoughNote is the cough trigger. At it's core, our cough trigger is a 20 kHz band-pass filter with a Q-factor of 20, which feeds into a manually tuned comparator. When the comparator is activated, it wakes up the system microcontroller, which proceeds to store a 1 s audio clip. Although the cough trigger is conceptually simple, it is carefully designed to activate on coughs, not activate on speech, and not be overly sensitive to environmental noise.

Cough vs Voice

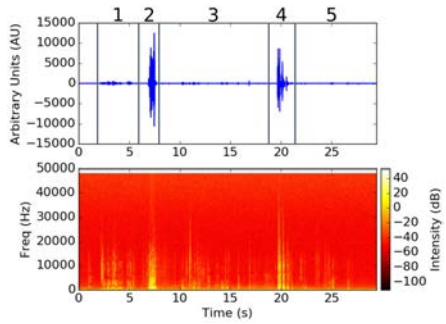
The primary goal of CoughNote's cough trigger is to alert the microcontroller to the potential start of a cough, while also not triggering on voiced audio, even if the voiced audio is high volume. There are three reasons for this. First, we expect that most sensitive audio consists of speech or other vocalized sounds. Second, other cough counters have reported speech as the dominant source of false positives [119, 123]. Third, we expect that in many environments, background talking is the dominant source of environmental noise. Thus, the most important consideration of the cough trigger is to distinguish between the audio characteristics of coughs and vocalized audio.

Coughs have 2-3 audible phases: explosive, intermediate, and (sometimes) voiced, which be seen in Figure 5.3a [149]. The explosive phase is typically the highest amplitude phase and gives coughs their characteristic sound [133]. Furthermore, the explosive phase is similar across individuals in both the time and frequency domain, and even across healthy patients vs patients with chronic bronchitis [148]. In contrast, various diseases can cause significant changes in the intermediate phase [146, 150]. For these reasons, most cough detection systems focus on detecting the explosive phase of a cough.

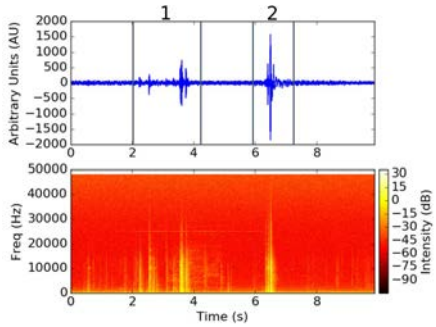
In addition to being the loudest part of a cough, the explosive phase generates a wide frequency band of sound. It is well established that coughs generate noticeable sound up to at least 20 kHz [146], with some studies indicating that coughs may generate noticeable sound up to 80 or 90 kHz [151]. Furthermore, studies have found noticeable 20 kHz sound



(a) Example Cough



(b) Comparison of Talking, Yelling, and Coughing



(c) Door Opening and Closing

Figure 5.3: 5.3a shows an example cough from a healthy, non-smoking individual. This cough shows all three cough phases: explosive (1), intermediate (2), and voiced (3, optional). 5.3b shows relative amplitudes and spectrograms of conversational voice (1,3), loud, high-pitched yelling (2), and a series of coughs (4). Harmonics from a normal conversational voice can reach 10 kHz, while high-pitched yelling can reach up to 15 kHz. Even high-pitched yelling does not present significant energy at 20 kHz, while the coughs contain significant energy up to 30 kHz. 5.3c shows a door shutting, which generates a high amplitude, wide-band sound, which can trigger a false positive recording.

in voluntary [150], chemically induced [147], and spontaneous coughs [151], indicating that explosive coughs likely generate high-frequency sound for a wide variety of cough causes. Asthmatic coughs may be an exception to this, since they may have a different sound generating mechanism than other coughs [152]. Our own testing on voluntary coughs found that coughs consistently produce sound past 20 kHz and often produce sound up to 45 kHz, the limit of our testing equipment, as seen in Figure 5.3a.

In contrast, voices produce a much lower range of frequencies. The fundamental frequency is typically around 200 Hz for adult females and 125 Hz for adult males [153], meaning that we would expect to see very little noise at even 20 kHz. Although it is difficult to directly compare this to coughs, which are not well characterized by a single fundamental frequency [154], the key idea here is that coughs produce much more high frequency noise,

Audio Label	Energy Above 5 kHz	Energy Above 10 kHz	Number of Audio clips	Source
Cough	3.5%	0.35%	61	AudioSet
Speech	1.8%	0.09%	200	AudioSet
Laughter	2.7%	0.14%	119	AudioSet
Vehicles	1.1%	0.005%	200	AudioSet
Music	2.0%	0.16%	200	AudioSet
Gun shot	10.2%	2.3%	374	Urban Sound 8k
Air Conditioner	8.6%	2.0%	1000	Urban Sound 8k
Drilling	22.6%	7.8%	1000	Urban Sound 8k
Jackhammer	18.1%	4.9%	1000	Urban Sound 8k
Children Playing	1.9%	0.37%	1000	Urban Sound 8k
Car Horn	1.6%	0.37%	429	Urban Sound 8k
Engine Idling	3.1%	0.91%	1000	Urban Sound 8k
Dog Barking	2.8%	1.0%	999	Urban Sound 8k
Street Music	1.3%	0.29%	1000	Urban Sound 8k
Siren	1.1%	0.15%	929	Urban Sound 8k

Table 5.1: Averages percentage of total audio energy found in high-frequency bands from categories in Google Audioset and the Urban Sound 8k DataSet. We selected AudioSet categories that we believe are common in everyday life, while the Urban Sound 8k DataSet is specifically tailed to sounds that commonly cause noise complaints in urban areas.

even when the vocal sound is much higher in amplitude [123], as illustrated in Figure 5.3b. Based on these differences between coughs and voiced sounds, the amplitude of sound above 20 kHz can effectively distinguish between coughs and vocalized sounds.

Environmental Noise

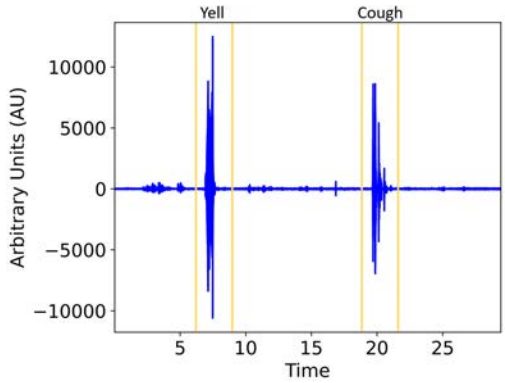
In addition to comparing coughs and voice, we investigated other sources of environmental noise to determine what other types of sounds produce 20+ kHz noise. To explore our environmental sound characteristics, we made a series of recordings in an embedded systems lab, a downtown coffee shop, and walking around the downtown area of the nearest city. These recordings were made using a TASCAM DR-05 recorder which has a relatively flat frequency response from 20 Hz to 40 kHz [155], at a sampling rate of 96 kHz to take full advantage of the recorder's frequency range. We then listened to these recordings (around 3.8 hours total), and examined spectrograms of the audio files. We found that ambient noise was generally low-frequency, with noise above even 10 kHz being uncommon, and in these settings the ambient noise was largely dominated by people talking. Ambient noise was very low in amplitude in comparison to recordings of a cough made with the same recorder held at the coughing person's neck. For reference, ambient talking in a crowded coffee shop was consistently under 20% of the maximum amplitude of the cough recording. We did find that sharp, metallic sounds, such as an electric drill or metal framed door slamming consistently generate high-frequency and high-amplitude noise, as can be seen in Figure 5.3c. In general, we found that events that generated 20+ kHz noise generated wide band noise, much like a cough, rather than specifically high frequency noise.

To further explore the effects of environmental noise, we also examine pre-existing audio

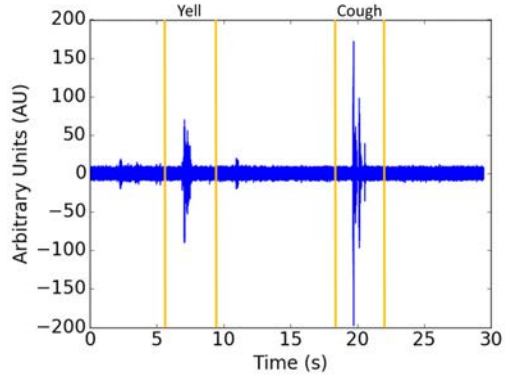
data sets. To the best of our knowledge, there is no publicly available data set that captures 20+ kHz sound. Typical consumer microphones, and even high-end vocal microphones, often have a frequency cutoff at around 10 kHz, since frequncies beyond this do not matter much for vocal audio [156, 157]. Instead of directly looking for 20 kHz noise, we examine the Google Audioset and Urban Sound 8k data sets for sound categories that produce relatively wide band sound between 0 and 10 kHz. Google Audioset is sourced from over 2 million YouTube videos, and contains an expansive array of labeled human, animal, and mechanical sounds, including coughs [158]. The Urban Sound 8k data set is sourced from FreeSound.org and is targeted towards sounds that commonly generate noise complaints in urban areas [159]. We analyzed the entire Urban Sound 8k data set, and selected categories from Audioset that we believed contained common everyday sounds not found in the Urban Sound 8k data set. Audioset clips are drawn from the "evaluation" section in each category, and we analyze up to 200 clips per category depending on availability.

To analyze the data sets, we took power spectral densities of each audio clip, and calculated the average energy contained above 5 kHz and above 10 kHz for each category, with the expectation that wide-band sound that could prove problematic would contain a higher proportion of their energy above 5 kZ and 10 kHz compared to narrow-band sound such as talking. The results of this analysis are shown in Table 5.1. As expected, vocal sounds such as speech, laughter, and music, contained a lower proportion of their energy above 5 kHz than coughs. Other sources of urban noise that are targeted at humans, such as sirens and car horns, also contain a lower proportion of their energy above 5 kHz. Mechanical, metallic sources of sounds such as drills and jackhammers contain significantly more of their energy above 5 kHz compared to coughs. In most circumstances, we would expect users not to spend too much time near these sources of noise. The only potentially problematic source of sound we found is air conditioner units, which contain significantly more of their energy above 5 kHz than coughs, and are commonly used in indoor environments. The sound clips of air conditioners appear to be recorded very close to air conditioner units, and while the ac unit is actively cooling. We found no such sound profile in our recordings of centrally cooled rooms, and in later testing centrally cooled rooms did not activate our cough trigger.

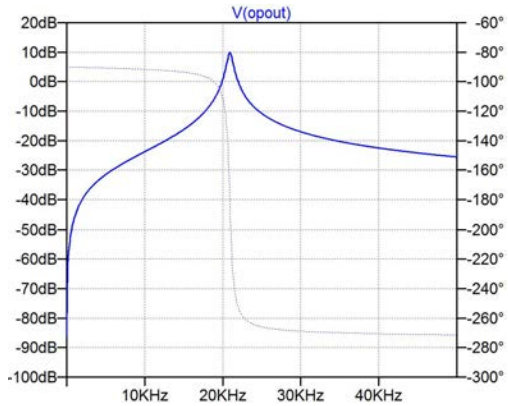
This analysis also does not take into account the sound intensity perceived by CoughNote's microphone. Sound intensity has an inverse squared relationship with distance, so doubling the distance between a source and a receiver halves the received sound intensity. In our system, the microphone is worn very near the neck, so the coughs are recorded at extremely close distances. Coughs are typically between 70-90 decibels [127, 160], or about as loud as heavy city traffic. Assuming the microphone is about 15 cm away from the cough, a sound identically loud as a cough but 2 m from the CoughNote user would be 22.5 dB lower form the perspective of the microphone. In addition, sounds between 80-85 decibels an cause hearing damage in under two hours, so we do not expect most people to spend significant time around similarly loud sounds [161]. These factors, combined with the paucity of high frequency sound found in our recordings and the analyzed data sets, show that a threshold-based trigger of sound above 20 kHz would not only be able to distinguish between coughs and voice, but also be resilient to environmental noise in everyday situations.



(a) Unfiltered yelling vs coughing



(b) Filtered yelling vs coughing



(c) Audio filter frequency response

Figure 5.4: Figure 5.4a shows us an audio waveform of a yell followed by a cough. Figure 5.4b shows us the approximate effect of our filter (Figure 5.4c) on Figure 5.4a. The unfiltered yell is higher amplitude than the unfiltered cough, but the filtered cough is over 3x the amplitude of the filtered yell. Figure 5.4c shows the simulated frequency response of the filter used in our cough trigger. Our filter is theoretically centered at 20 kHz, and empirically centered at 18.6 kHz. The bright blue curve shows the attenuation of our filter, while the faded curve behind it is the phase response of the filter. For this application, the phase response is relatively unimportant.

Cough Trigger Design Details

Based on our analysis of coughs, voice, and everyday environmental noise, we implemented our cough trigger as a multiple-feedback band-pass filter theoretically centered around 20 kHz, which feeds into a comparator. When a manually tuned comparator threshold is hit, the comparator generates an interrupt to the system microcontroller. The simulated frequency response of our filter can be seen in Figure 5.4c. Empirically, we found the filter to be centered around 18.6 kHz. We chose to use a band-pass filter over a high-pass filter because our selected microphone has a relatively low attention, -4 dB, even at 80 kHz, and we are unable

to evaluate sound above 46 kHz with our current implementation. Out of an abundance of caution, we wanted to ensure CoughNote was not affected by environmental noise above 46 kHz that we could not previously notice. We chose 20 kHz as our target center frequency because sound in at around 20 kHz was already very rare in our recordings, so even though our recorded coughs generated sound beyond that, there did not appear to be much benefit to setting the center frequency higher. Because the filter dramatically attenuates the amplitude of the cough signal, we also add a 10 dB gain to the filter to make the output easier to work with.

Figure 5.4b shows the simulated effects of our filter on the yell and cough seen in Figure 5.4a. Although the yell is higher in amplitude than the cough, the filtered cough has over three times the maximum amplitude of the filtered yell. In addition, the yell was given at the highest pitch the participant could produce, with a fundamental frequency of 590 Hz, far higher than the typical fundamental frequency of adult speech (125-200 Hz) [153]. To calibrate the comparator threshold, we calibrate against explosive, or loud, voluntary coughs from 4 graduate students, and manually tuned the threshold to activate at the beginning of their coughs. It was trivial to set a threshold that consistently activated on their coughs, but did not active on their speech or yelling, loud music from a speaker, or common soft mechanical sounds such as a mechanical keyboard.

5.4 Implementation

Here, we wish to explain our fast-storage layer in more detail, and provide some guidance on the parts we used.

Fast Storage Layer

Since CoughNote's cough trigger uses sound around 20 kHz, we want our system to record audio with a sampling rate of at least 40 kHz. In addition, recent cough sensors utilize 16 bit audio [122, 123], so we follow suit with CoughNote. A 1 s audio clip thus requires about 80 kB of memory. Since standard Cortex-M microcontrollers typically contain between 16-64 kB of RAM, the audio clip must be offloaded to flash during recording. Without using DMA, we found that using the MCU as an intermediary between an external adc and an external flash chip resulted in a 20 kHz sampling even high transfer clock speeds. Unfortunately, using DMA added a startup lag to our recording process, causing us to miss part of the explosive phase of the cough. Specialized sound recorder ICs which consume 10s of mWs and have enough memory exist, but these are targeted at recording voices, and rarely offer a 40 kHz sampling rate.

To enable audio capture without using a higher powered class of microcontroller or highly tuned processor selection and driver code, we designed a fast storage layer, shown in Figure 5.5. This layer uses two digital switches and a SPI mode compatible FRAM chip and ADC, and enables any microcontroller to directly save audio samples to the FRAM chip, bypassing

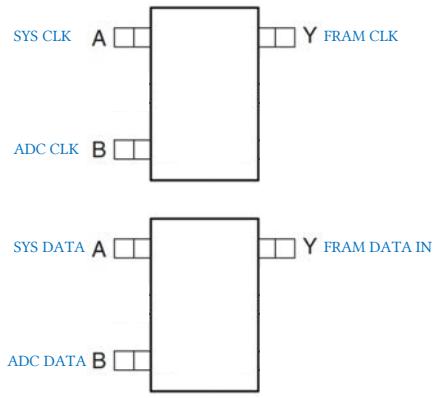


Figure 5.5: Fast storage layer consisting of two digital switches, an ADC, and an FRAM chip. When set to A, the switches allow the mcu to directly communicate with the FRAM chip. When set to B, the switches short the ADC's data output line to the FRAM's data input line, and shorts the two clock lines together. The MCU first uses the A setting to initialize the FRAM chip. It then uses the B setting to short both clock lines together, and by running the clock line directly stores the ADC output to the FRAM chip.

the MCU entirely. Initially, the FRAM chip and ADC are connected separately to the microcontroller, which allows the microcontroller to set up each chip. Then, the switches short the FRAM chip and ADC clock lines together with the microcontroller provided SPI clock line, and connect the ADC data out and FRAM data in lines. In this way, the ADC is able to directly write its data to the FRAM chip, bypassing the RAM limitations of the microcontroller. Although an FRAM chip is not strictly necessary, the memory chip used must either not have a page system or have pages larger than the incoming data. Otherwise, the time delay caused by stopping the ADC and switching memory pages will result in a lower and uneven sampling rate. This doubles our implementation's sample rate, allowing it to consistently sample the audio signal at 40 kHz.

Parts

CoughNote is constructed entirely from COTS parts, although care must be taken when selecting certain parts. In particular, the microphone should frequency cut-off of at least 20 kHz, an unusual requirement for many MEMS microphones. We selected the SPU0410LR5H-QB for our microphone because at our cough trigger's target center frequency (20 kHz), the SPU0410LR5H-QB provides a 8 dB gain, making it easier to work with frequencies in that area. For our band-pass filter, we use the MAX9912 op-amp, and to buffer our signals, we use the MIC863 op-amp. The choice of buffer op-amp is flexible, but we caution on replacing the MAX9912 with care. In limited testing, we found that the filter circuit could easily become unstable with other low-powered op-amps. We used a microSD card for our permanent storage layer for convenience. To keep sleep current low, we disconnected the card from VCC

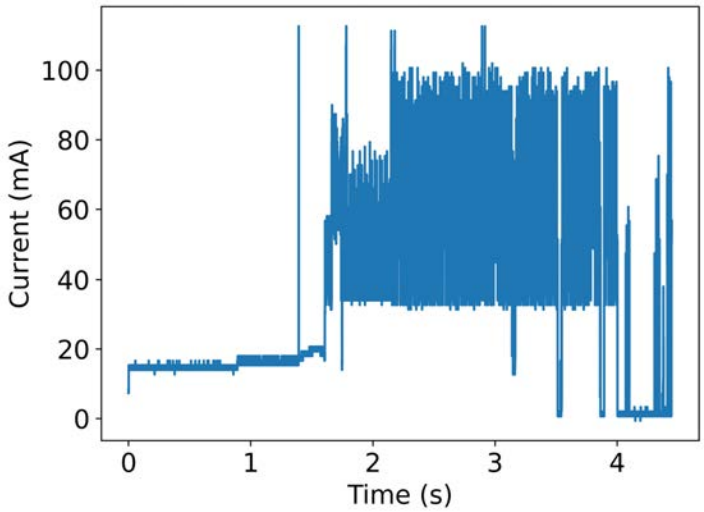


Figure 5.6: Current trace for recording a 1 s audio clip

when not in use. However, the microSD card's long start-up process combined with the high overhead of sending many small chunks of data rate limited CoughNote's ability to record audio to once every 5 seconds. It further imposed a high energy cost, which is discussed in Section 5.5. Finally, our full hardware and software stack can be found on GitHub.

5.5 Evaluation

The goal of CoughNote is not to obtain exact cough counts. Rather, CoughNote seeks to capture a sufficient number of coughs to detect general trends in coughing, in support of determining objective onset and disappearance times of coughs. The primary goal of CoughNote then is to be able to record cough trends. The more frequently someone coughs, the higher cough rate CoughNote should record. In our evaluation, we establish baseline power and performance metrics in controlled tests and a small pilot study, and then evaluate how well CoughNote captures cough trends and generalizes to different people and causes of cough by joining a pre-existing challenge study involving 4 influenza patients. Both the pilot study and challenge study were reviewed and approved of by the University of Michigan Institutional Review Board.

Evaluating Energy Use

CoughNote's power usage can be simply modeled using sleep power and the energy required to record a 1 s audio clip. A current trace for recording a 1 s audio clip upon the cough trigger activating is shown in Figure 5.6. In total, recording a 1 s audio clip requires 516

mJ. 422 out of the 516 mJ is consumed by powering on the microSD card, performing the necessary setup, and writing the data to it, which is represented by the high current draw from 2 s onward in Figure 5.6. Theoretically, the sleep current for CoughNote should be 156 μA. This is dominated by our microphone, which consumes 120 μA at all times. The cough-trigger circuitry itself consumes only 14 μA. In our prototype implementation, CoughNote empirically has a sleep current of 523 μA due to unoptimized processor sleep code. Empirically, idle current draw consumes 621 mJ per 24 hours, but with optimized sleep code we expect CoughNote's idle current draw to consume 185 mJ per 24 hours.

Prior work suggests that cough monitors should last at least a day [116]. CoughNote uses a 110 mAh LiPo battery, which provides 1,465 J. Subtracting out 621 mJ for sleep leaves CoughNote with enough energy for 2837 recordings over 24 hours. Based on our pilot study, a more realistic estimate would be 10 false positives per hour plus coughs. Based on prior work, bronchitis patients on average cough 213 times a day [126]. Using bronchitis patients as a rough model for sick users, we would expect CoughNote to consume 235 J per day, giving it a battery life of 6.2 days. The storage operations account for over 99% of the energy budget, even though we are only recording 7.5 minutes of audio per day. This is largely due to the power requirements of using a microSD card, which have high current draws and impose a high overhead for repeated small writes. For uses that require a longer battery life and do not require the convenience of a microSD card, power draw can be drastically reduced by replacing the microSD card with a NAND flash chip.

Even with a unoptimized permanent storage layer and unoptimized sleep code, the battery life of CoughNote still outperforms systems which require users to constantly record audio, while being smaller and lighter. A top rated voice recorder on Amazon has an advertised battery life of 57 hours, weighs 74 g, and measures 1.51 x 4.49 x 0.76 in [162]. In comparison, CoughNote provides a 148 hour battery life, weighs 15.9 g, and measures 1.5 x 1.6 x 0.6 in.

Pilot Test

To establish baseline performance metrics, we conducted a pilot test involving one non-coughing and one-coughing user. The non-coughing user wore CoughNote for a total of 112 hours. The coughing user wore CoughNote for a total of 53 hours, and logged timestamps for 23 cough episodes. A cough episode consists of all coughs that occur together. The coughing user did not record the number of coughs per episode. Combined, the two participants wore CoughNote in a wide variety of environments, including homes, restaurants, coffee shops, and an embedded systems lab. After the deployment, we collected the data, and manually listened to the audio clips.

We found that CoughNote successfully recorded 15 out of the 23 cough episodes (65%). In addition, the non-coughing user's sensor captured 5.03 false positives an hour, while the coughing user's sensor captured 15 false positives an hour. While the non-coughing user's false positives sounded loud and mechanical, and contained no discernible voices, many of the coughing user's false positives were quiet. More alarmingly, some of these quiet false positives even contained an intelligible word, although none of the false positives appeared to

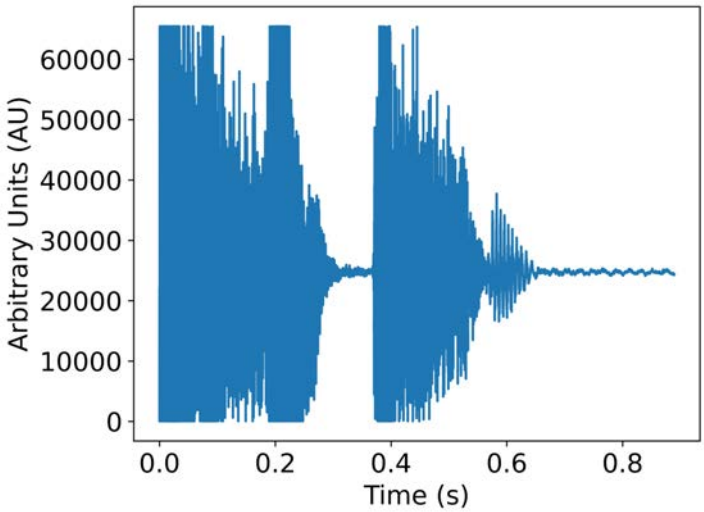


Figure 5.7: A typical cough waveform from our wearable device. In this particular recording, two short coughs are captured.

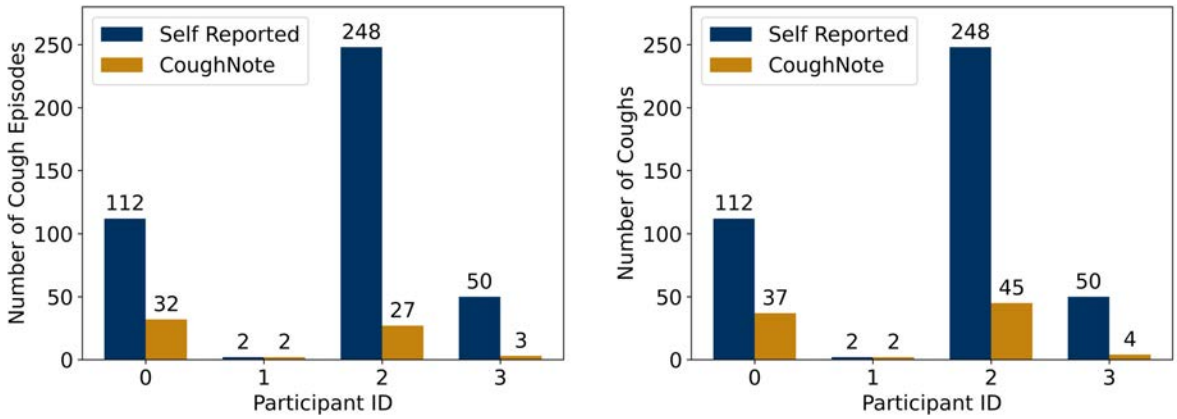
be triggered by a word. Upon further examination and testing, we found that the coughing user wore a button down shirt and wore the CoughNote sensor near a button. We found that rubbing CoughNote's case against the button could easily trigger a quiet false positive. In contrast, the non-coughing user wore a smooth shirt, and the case rubbing against fabric did not produce any false positives.

Qualitatively, the audio quality is clear, and coughs are easy to identify. An example of two captured coughs is shown in Figure 5.7. Particularly loud coughs, such as those shown in Figure 5.7, can experience minor clipping. Although the cough audio is qualitatively still clear, CoughNote would benefit from more finely tuned gain control or dynamic gain control. In addition, when users directly cough on the mic, CoughNote can capture the characteristic "pop" that results from fast-moving air over a mic, and would benefit from a pop filter. Because the trigger occurs after the start of the cough, we estimate that we lose approximately the first 10 ms of a cough, which in Figure 5.7 can be seen as the slight rise in the second cough that is missing in the first cough. Subjectively, this does not affect the audio clarity of the cough, and we found no evidence that this information loss matters.

Overall, our pilot study establishes a baseline cough episode recording rate of 65%, an average of 10 false positive recordings an hour, and that voices do not active our cough trigger.

Challenge Study

To evaluate how well CoughNote generalizes, we joined a pre-existing challenge study. Broadly speaking, a challenge study involves participants getting purposely infected, or challenged,

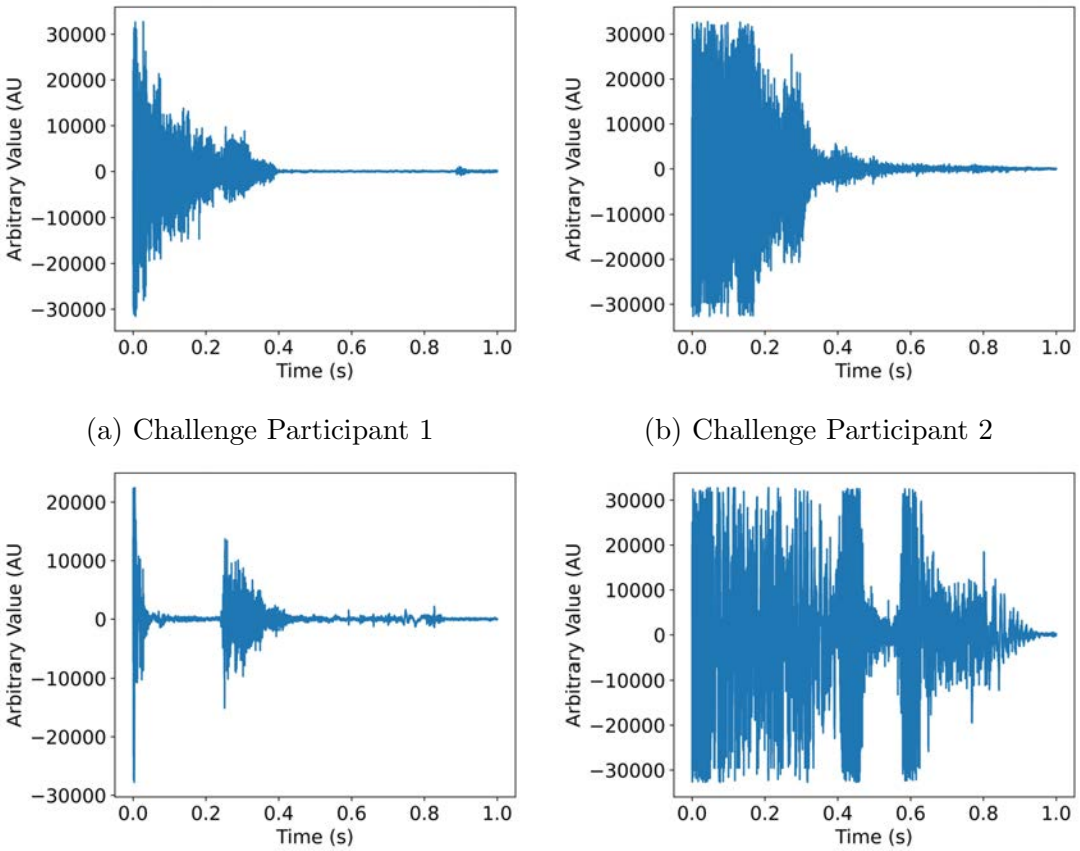


(a) Cough Episodes (b) Cough Counts

Figure 5.8: Self-reported cough estimates vs CoughNote recorded coughs. In 5.8a, we count the number of audio-clips that a participant's CoughNote recorded that contains at least one cough. In 5.8b, we count the total number of coughs captured in CoughNote's recorded clips. Because participants only estimated a "number of coughs", their self-reported count is the same for both metrics. Participant cough counts are not based on detailed logs or diaries, but just represent the participants' best guess as to how much they coughed in a day. In addition, while users estimated their total cough counts for a day, they often did not necessarily wear CoughNote for the whole day. Participants wore CoughNote for a varying number of hours and a varying number of days. These numbers are aggregated over the entire time users wore CoughNote. These results suggest that CoughNote can successfully capture cough count trends, especially when we take into account multiple coughs being present in a single audio snippet.

with a disease. The effects of the disease or therapeutic measures can then be studied in detail. The challenge study we joined involved participants who were infected with a low-symptom strain of influenza, and kept in a hospital for a week. Multiple institutions were involved in this study, and we joined in conjunction with epidemiologists from the University of Michigan, Ann Arbor.

Our goal with this test is to evaluate CoughNote's ability to generalize to other people, and other causes of coughs. 15 participants in the challenge study wore CoughNotes during the study for a total of 185 hours. After the end of each day they wore the sensor, the participants reported an estimated number of coughs that day. In total, we collected self-reported estimates and CoughNote recorded coughs from four participants. Three other participants reported between 1-3 coughs but had no CoughNote recorded coughs. The other eight participants recorded no coughs and reported no coughs. Qualitatively, audio was clear, coughs were easy to identify, and we did not notice any intelligible words. On average, each participant's CoughNote recorded 10.3 false positives per hour and 0.66 coughs per hour. Even with this



(a) Challenge Participant 1

(b) Challenge Participant 2

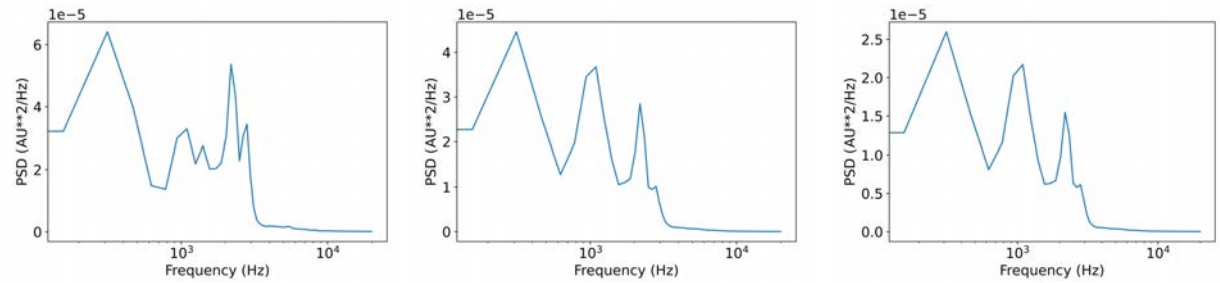
(c) Challenge Participant 3

(d) Challenge Participant 4

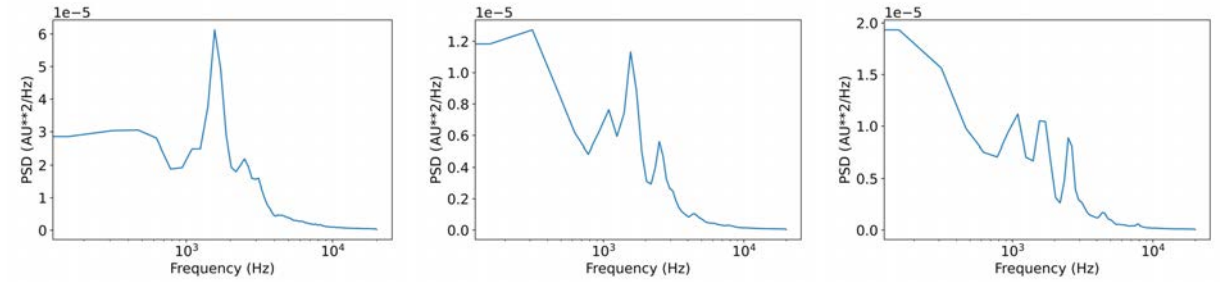
Figure 5.9: Example cough waveforms from challenge study participants with recorded coughs. Subjectively, we found that cough sounds varied between coughs of an individual, and especially between coughs from different individuals. These examples illustrate the diverse array of coughs recorded by CoughNote.

high false positive rate, CoughNote still records over 300x less data than an always-on audio recorder. Furthermore, we found no vocalized or otherwise sensitive audio in CoughNote's recordings. Because we lack detailed cough logs or other reliable forms of ground truth, we cannot directly evaluate what proportion of coughs CoughNote records. With the key assumption that self-reported cough count estimates are correlated with true cough counts, we instead evaluate how well CoughNote correlates with self-reported cough count estimates.

A summary of the results of the four participants with both self-reported and sensor recorded coughs is shown in Figure 5.8. It is unclear if participants attempted to estimate the number of cough episodes or their actual total number of coughs. We have included both metrics from CoughNote in Figure 5.8. CoughNote's cough episodes have a strong rank correlation (r=0.80), while CoughNote's cough counts have a perfect rank correlation



(a) Pilot Cough PSD - 100 ms (b) Pilot Cough PSD - 500 ms (c) Pilot Cough PSD - 1000 ms



(d) Challenge Cough PSD - 100 ms (e) Challenge Cough PSD - 500 ms (f) Challenge Cough PSD - 1000 ms

Figure 5.10: Average cough power spectral densities (PSD) for the pilot and challenge studies. The first 100 ms of a cough roughly represents the explosive phase of the cough, while the first 500 ms represents the total duration of the first cough in an audio clip. Although there are sometimes follow-up coughs or other noise in the last 500 ms of the audio, we found that virtually all of the energy is contained in the first 500 ms of the audio clip. In both the pilot study

correlation with self-reported cough estimates. We believe this shows that CoughNote can detect trends in coughing, and can work well on a variety people, not just the people we calibrated and piloted the sensor with.

To further evaluate how well CoughNote generalizes, we compare the captured coughs between our pilot study and the challenge study. Qualitatively, the coughs in the challenge study sounded more varied than the coughs in the pilot study. While coughs in the pilot study sounded roughly equal in pitch and and sharpness, coughs in the challenge study sometimes sounded very coarse and other times very sharp, and varied in pitch. Figure 5.9 illustrates some of the diverse array of coughs CoughNote recorded for the challenge study. Coughs are well known to vary in sound based on the cause. Although the cause of the pilot study participant's cough is unknown, it was likely not caused by influenza given the lack of other symptoms and time of year. Figure 5.10 compares the average power spectral densities of coughs in the pilot study and coughs in the challenge study, which are likely caused by the influenza injected into the participants. We find that for both pilot and challenge participants,

coughs almost universally range from 200-500 ms in duration. Furthermore, as Figure 5.10e and Figure 5.10b show, both sets of coughs demonstrate frequency peaks at around the same frequencies, and both show a triple peak structure. However, we found a significant difference in the structure of the explosive phase of the coughs, which is the portion of the cough that CoughNote's cough trigger rlies on. These differences can be seen in Figure 5.10a and Figure 5.10d, with the pilot study cough showing a double peak structure, while the challenge study showed only a single peak. In addition, the largest peak Figure 5.10a does not match the frequency of the single peak in Figure 5.10d. Despite these differences, CoughNote was still able to capture coughs in both sets of participants.

Between the diverse array of coughs captured and the strong correlation between CoughNote cough counts and self-reported cough count estimates, the pilot study and challenge study provide strong evidence that CoughNote performs well enough to track general trends in cough counts and generalizes across people and cough causes. CoughNote accomplishes this while offering a smaller form factor, higher-battery life, and minimal privacy intrusions compared to previous systems which require users to constantly record their audio environment. Having demonstrated the viability of a small and privacy preserving design, we leave to future work and improving technology the task of further reducing power, increasing cough recording rates, and increasing the recorded audio quality.

5.6 Summary

While research on creating algorithms to accurately count coughs in an audio file continues to advance the state of the art, we find that in many cases, simply knowing cough trends would be incredibly valuable, and that the value of precise cough counts is often murky. While continuous audio recording ensures that all coughs are captured for processing, in many applications, such as determining symptom onset, we found it was hard to justify the massive privacy implications of doing so. To alleviate this problem, we prioritize the user over the system, designing a wearable sensor that prioritizes not capturing sensitive spoken audio over recording every single cough. CoughNote does this by utilizing passive vigilance in the audio domain to capture potential coughs while avoiding private spoken audio. This is enabled by a novel hardware cough trigger that can distinguish cough-like sounds from speech, and is smaller, lighter, and has three times the battery life of a typical voice recorder. While CoughNote does not capture every cough and may indeed fail to capture many coughs in the wild, we believe that focusing on privacy and usability is an important step forward for the practical deployability of cough sensors.

Chapter 6

Conclusion

Wearable sensors offer a tantalizing method of studying and interacting with human health in ways never before possible. Not only does academic research continue to push the limits of wearable sensing applications, but we are beginning to see significant advancements in commercially available wearable sensors. Products like the Apple Smart Watch can now perform ECGs and measure blood oxygen levels, and the Empatica Embrace wrist band can be used to detect oncoming seizures. Many of these advancements have been enabled and driven by advancements in micro-processors, radios, and sensor ICs. Better sensor ICs have given wearables the ability to conveniently access a wide variety of data, and advancements in micro-processors and radios have allowed researchers to increasingly treat wearables like normal computers. Modern radios allow wearables to easily interact with users by tethering to smartphones, and modern micro-processors enable local data analysis and machine learning on a wearable [163, 164].

However, even with these advancements, wearables will continue to be limited by battery capacity. Smart phones were able to adapt to increasingly powerful hardware in part by significantly increasing in size and battery capacity [165]. So far, there is no evidence that users are willing to embrace that same trend in wearables. This constraint is only exacerbated by interest in alternative wearable sensor form factors such as rings and temporary tattoos [86, 166]. This suggests that there will continue to be significant value in looking for sensing methodologies outside of the traditional sample-process-store framework for wearable sensing. In this dissertation, we have demonstrated that by utilizing passive-vigilance instead of active data processing in the ultrasonic, accelerometric, and acoustic domains, we can build novel and usable wearable sensors for a variety of applications. We note that the work in this dissertation is not a product of technology that was simply unavailable to prior system designers. The only critical hardware component to Opo's design was released in 2001, 13 years before its initial publication [4]. Similarly, our hand-washing and cough sensing systems were not enabled by newly available components. These works were all enabled by novel architectures to realize passive vigilance, that is, sensing without engaging the processor or radio. We hope this dissertation inspires future system designers to do the same, enabling the construction of domain-specific toolkits for a range of societally valuable applications.

Bibliography

[1] Shweta Bansal et al. "Big data for infectious disease surveillance and modeling". In: *The Journal of infectious diseases* 214 (2016), S375–S379.

[2] Robert Moss et al. "What can urban mobility data reveal about the spatial distribution of infection in a single city?" In: *BMC public health* 19.1 (2019), pp. 1–16.

[3] Timo Smieszek. "A mechanistic model of infection: why duration and intensity of contacts should be included in models of disease spread". In: *Theoretical Biology and Medical Modelling* 6.1 (2009), p. 25.

[4] William Huang et al. "Opo: A Wearable Sensor for Capturing High-Fidelity Face-to-Face Interactions". In: *Proceedings of the 12th ACM Conference on Embedded Network Sensor Systems*. ACM. 2014, pp. 61–75.

[5] Paul Zivich et al. "Measuring office workplace interactions and hand hygiene behaviors through electronic sensors: a feasibility study". In: (2021).

[6] Gary W Evans and Richard E Wener. "Crowding and Personal Space Invasion on the Train: Please Don't Make Me Sit in the Middle". In: *Journal of Environmental Psychology* 27.1 (2007), pp. 90–94.

[7] Cade McCall et al. "Proxemic Behaviors as Predictors of Aggression Towards Black (but not White) Males in an Immersive Virtual Environment". In: *Social Influence* 4.2 (2009), pp. 138–154.

[8] Lynn Wu et al. "Mining face-to-face interaction networks using sociometric badges: Predicting productivity in an it configuration task". In: *Available at SSRN 1130251* (2008).

[9] Elena Stephan, Nira Liberman, and Yaacov Trope. "Politeness and Psychological Distance: a Construal Level Perspective." In: *Journal of personality and social psychology* 98.2 (2010), p. 268.

[10] Catherine Beaulieu. "Intercultural Study of Personal Space: A Case Study". In: *Journal of applied social psychology* 34.4 (2004), pp. 794–805.

[11] Nan M Sussman and Howard M Rosenfeld. "Influence of Culture, Language, and Sex on Conversational Distance." In: *Journal of Personality and Social Psychology* 42.1 (1982), p. 66.

[12] *Scientific Brief: SARS-CoV-2 and Potential Airborne Transmission*. 2020. URL: https://www.cdc.gov/coronavirus/2019-ncov/more/scientific-brief-sars-cov-2.html#:~:text=Droplet%5C%20transmission%5C%20is%5C%20infection%5C%20spread,generally%5C%20within%5C%20about%5C%206%5C%20feet..

[13] *Limiting Spread: Limiting the Spread of Pandemic, Zoonotic, and Seasonal Epidemic Influenza*. 2010. URL: http://www.who.int/influenza/resources/research/research_agenda_influenza_stream_2_limiting_spread.pdf.

[14] E Rea et al. "Duration and Distance of Exposure are Important Predictors of Transmission among Community Contacts of Ontario SARS Cases". In: *Epidemiology and infection* 135.06 (2007), pp. 914–921.

[15] Tze-Wai Wong et al. "Cluster of SARS Among Medical Students Exposed to Single Patient, Hong Kong". In: *Emerging infectious diseases* 10.2 (2004), p. 269.

[16] Centers for Disease Control and Prevention. *Social Distancing*. 2020. URL: https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/social-distancing.html.

[17] Thang Hoang et al. "A systematic review of social contact surveys to inform transmission models of close-contact infections". In: *Epidemiology (Cambridge, Mass.)* 30.5 (2019), p. 723.

[18] Timo Smieszek et al. "Contact diaries versus wearable proximity sensors in measuring contact patterns at a conference: method comparison and participants' attitudes". In: *BMC infectious diseases* 16.1 (2016), p. 341.

[19] P. Beutels et al. "Social Mixing Patterns for Transmission Models of Close Contact Infections". In: *Epidemiology & Infection* 134.6 (Dec. 2006), pp. 1158–1166.

[20] Siobhan K. Halloran et al. "A Comprehensive Breath Plume Model for Disease Transmission via Expiratory Aerosols". In: *PLoS One* 7.5 (May 2012).

[21] Anice C Lowen et al. "The Guinea Pig as a Transmission Model for Human Influenza Viruses". In: *Proceedings of the National Academic of Sciences of the United States of America* 103.26 (June 2006).

[22] Erik M. Volz Allison E. Aiello. *eX-FLU: A Social Network Intervention For Reducting Respiratory Infectious Illness Transmission*. http://www.screencast.com/t/pBxndEk65m. 2013.

[23] Wen Dong, Katherine Heller, and Alex Sandy Pentland. "Modeling infection with multi-agent dynamics". In: *Social Computing, Behavioral-Cultural Modeling and Prediction*. Springer, 2012, pp. 172–179.

[24] Youngki Lee et al. "Comon: Cooperative Ambience Monitoring Platform with Continuity and Benefit Awareness". In: *Proceedings of the 10th international conference on Mobile systems, applications, and services*. ACM. 2012, pp. 43–56.

[25] *Exposure Notification - Bluetooth Notification*. 2020. URL: https://covid19-static.cdn-apple.com/applications/covid19/current/static/contact-tracing/pdf/ExposureNotification-BluetoothSpecificationv1.2.pdf?1.

[26] Arkadiusz Stopczynski et al. "Measuring large-scale social networks with high resolution". In: *PloS one* 9.4 (2014), e95978.

[27] Eiko Yoneki and Jon Crowcroft. "EpiMap: Towards Quantifying Contact Networks for Understanding Epidemiology in Developing Countries". In: *Ad Hoc Networks* (2012).

[28] Andrzej Forys et al. "WRENMining: Large-Scale Data Collection for Human Contact Network Research". In: *Proceedings of First International Workshop on Sensing and Big Data Mining*. ACM. 2013, pp. 1–6.

[29] Moses Chapa Kiti et al. "Study design and protocol for investigating social network patterns in rural and urban schools and households in a coastal setting in Kenya using wearable proximity sensors". In: *Wellcome open research* 4 (2019).

[30] Mathieu Génois et al. "Data on face-to-face contacts in an office building suggests a low-cost vaccination strategy based on community linkers". In: *arXiv preprint arXiv:1409.7017* (2014).

[31] Marcel Salathé et al. "A High-Resolution Human Contact Network for Infectious Disease Transmission". In: *Proceedings of the National Academy of Sciences (PNAS)* 107.51 (2010), pp. 22020–22025.

[32] Douglas J Leith and Stephen Farrell. *Coronavirus Contact Tracing: Evaluating The Potential Of Using Bluetooth Received Signal Strength For Proximity Detection*. 2020. URL: https://www.scss.tcd.ie/Doug.Leith/pubs/bluetooth_rssi_study.pdf.

[33] W.F. Wells. "On Air-Borne Infection. Study II. Droplets and Droplet Nuclei". In: *American Journal of Hygiene* 20 (1934), pp. 611–618.

[34] Zixiang Ma et al. "A BLE RSSI ranking based indoor positioning system for generic smartphones". In: *2017 Wireless Telecommunications Symposium (WTS)*. IEEE. 2017, pp. 1–8.

[35] Douglas J Leith and Stephen Farrell. "Measurement-based evaluation of Google/Apple Exposure Notification API for proximity detection in a light-rail tram". In: *Plos one* 15.9 (2020), e0239943.

[36] Douglas J Leith and Stephen Farrell. "Measurement-based evaluation of google/apple exposure notification api for proximity detection in a commuter bus". In: *arXiv preprint arXiv:2006.08543* (2020).

[37] Thuraiappah Sathyan, David Humphrey, and Mark Hedley. "WASP: A System and Algorithms for Accurate Radio Localization Using Low-Cost Hardware". In: *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews* 41.2 (2011), pp. 211–222.

[38] Benjamin Kempke, Pat Pannuto, and Prabal Dutta. "Polypoint: Guiding indoor quadrotors with ultra-wideband localization". In: *Proceedings of the 2nd International Workshop on Hot Topics in Wireless*. 2015, pp. 16–20.

[39] Seyed Omid Sadjadi, Craig S Greenberg, and Douglas A Reynolds. "NIST Pilot Too Close for Too Long (TC4TL) Challenge Evaluation Plan". In: (2020).

[40] NIST. *TC4TL Challenge Leaderboard*. 2020. URL: https://tc4tlchallenge.nist.gov/.

[41] Andreas Biri et al. "SociTrack: Infrastructure-Free Interaction Tracking through Mobile Sensor Networks". In: *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*. MobiCom '20. London, United Kingdom: Association for Computing Machinery, Sept. 2020. ISBN: 9781450370851. DOI: 10.1145/3372224.3419190. URL: https://doi.org/10.1145/3372224.3419190.

[42] *TelosB Mote Platform*. http://www.willow.co.uk/TelosB_Datasheet.pdf.

[43] Nadav Aharony et al. "Social fMRI: Investigating and Shaping Social Mechanisms in the Real World". In: *Pervasive and Mobile Computing* 7.6 (2011), pp. 643–659.

[44] Nissanka B Priyantha, Anit Chakraborty, and Hari Balakrishnan. "The Cricket Location-Support System". In: *Proceedings of the 6th annual international conference on Mobile computing and networking (MobiCom)*. ACM. 2000.

[45] S. Park et al. "Design of a Wearable Sensor Badge for Smart Kindergarten". In: *Proceedings of the 6th International Symposium on Wearable Computers. (ISWC)*. IEEE. 2002, pp. 231–238.

[46] Masateru Minami et al. "DOLPHIN: A Practical Approach for Implementing a Fully Distributed Indoor Ultrasonic Positioning System". In: *UbiComp 2004: Ubiquitous Computing*. Springer, 2004, pp. 347–365.

[47] José Carlos Prieto et al. "Performance Evaluation of 3D-LOCUS Advanced Acoustic LPS". In: *Instrumentation and Measurement, IEEE Transactions on* 58.8 (2009), pp. 2385–2395.

[48] Brita Meriac Milosch Meriac. *The OpenBeacon project*. 2020. URL: https://www.openbeacon.org/.

[49] Steven Lanzisera and Kristofer S.J. Pister. "Burst Mode Two-Way Ranging with Cramer-Rao Bound Noise Performance". In: *Global Telecommunications Conference*. 2008.

[50] Pat Pannuto et al. "Harmonium: Ultra wideband pulse generation with bandstitched recovery for fast, accurate, and robust indoor localization". In: *ACM Transactions on Sensor Networks (TOSN)* 14.2 (2018), pp. 1–29.

[51] Chunyi Peng et al. "BeepBeep: A High Accuracy Acoustic Ranging System using COTS Mobile Devices". In: *Proceedings of the 5th international conference on Embedded networked sensor systems*. ACM. 2007, pp. 1–14.

[52] Patrick Lazik et al. "ALPS: A bluetooth and ultrasound platform for mapping and localization". In: *Proceedings of the 13th ACM conference on embedded networked sensor systems*. 2015, pp. 73–84.

[53] Kaikai Liu, Xinxin Liu, and Xiaolin Li. "Guoguo: Enabling Fine-grained Indoor Localization via Smartphone". In: *MobiSys '13: Proceeding of the 11th Annual International Conference on Mobile Systems, Applications, and Services*. 2013, pp. 235–248.

[54] Yasuhiro Fukuju et al. "DOLPHIN: An Autonomous Indoor Positioning System in Ubiquitous Computing Environment." In: *The Workshop on Software Technologies for Future Embedded and Ubiquitous Systems (SEUS)*. 2003.

[55] Andreas Savvides, Chih-Chieh Han, and Mani B Strivastava. "Dynamic Fine-Grained Localization in Ad-hoc Networks of Sensors". In: *Proceedings of the 7 th annual international conference on Mobile computing and networking (MobiCom)*. ACM. 2001, pp. 166–179.

[56] Charles E. Perkins and Pravin Bhagwat. "Highly Dynamic Destination-Sequenced Distance-Vector Routing (DSDV) for Mobile Computers". In: *Proceedings of the conference on Communications architectures, protocols and applications*. ACM. 1994.

[57] Lorenzo Isella et al. "Close Encounters in a Pediatric Ward: Measuring Face-to-Face Proximity and Mixing Patterns with Wearable Sensors". In: ().

[58] Andy Harter et al. "The Anatomy of a Context-Aware Application". In: *Wireless Networks* 8.2 (2002), pp. 187–197.

[59] Joseph Polastre, Jason Hill, and David Culler. "Versatile Low Power Media Access for Wireless Sensor Networks". In: *Proceedings of the 2nd ACM Conferences on Embedded Networked Sensor Systems*. ACM. 2004, pp. 95–107.

[60] Prabal Dutta et al. "Design and Evaluation of a Versatile and Efficient Receiver-Initiated Link Layer for Low-Power Wireless". In: *Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems (SenSys)*. ACM. 2010, pp. 1–14.

[61] Michael J. McGlynn and Steven A. Borbash. "Birthday Protocols for Low Energy Deployment and Flexible Neighbor Discovery in Ad Hoc Wireless Networks". In: *Proceedings of the 2nd ACM International Symposium on Mobile Ad hoc Networking & Computing*. ACM. 2001.

[62] Arvind Kandhalu, Karthik Lakshmanan, and Ragunathan Raj Rajkumar. "U-Connect: A Low-latency Energy-efficient Asynchronous Neighbor Discovery Protocol". In: *Proceedings of the 9th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*. ACM/IEEE. 2010, pp. 350–361.

[63] Christine Julien et al. "Blend: practical continuous neighbor discovery for bluetooth low energy". In: *2017 16th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*. IEEE. 2017, pp. 105–116.

[64] Nathan E. Roberts and David D. Wentzloff. "A 98 nW Wake-up Radio for Wireless Body Area Networks". In: *IEEE Radio Frequency Integrated Circuits Symposium (RFIC)*. June 2012, pp. 373–376.

[65] Emanuele Lattanzi et al. "A Sub-µA Ultrasonic Wake-Up Trigger with Addressing Capability for Wireless Sensor Nodes". In: *ISRN Sensor Networks* (2013).

[66] Kshitij Yadav, Ioannis Kymissis, and Peter R. Kinget. "A 4.4 µW Wake-Up Receiver Using Ultrasonic Data". In: *IEEE Journal of Solid-State Circuits* 48.3 (2013), pp. 649–660.

[67] Li Liu et al. "Short-range airborne transmission of expiratory droplets between two people". In: *Indoor Air* 27.2 (2017), pp. 452–462.

[68] Juliette Stehlé et al. "Simulation of an SEIR Infectious Disease Model on the Dynamic Contact Network of Conference Attendees". In: *BMC medicine* 9.1 (2011), p. 87.

[69] Howard W Stoudt. *Skinfolds, Body Girths, Biacromial Diameter, and Selected Anthropometric Indices of Adults, United States, 1960-1962*. 1970.

[70] G. Oberholzer, P. Sommer, and R. Wattenhofer. "SpiderBat: Augmenting Wireless Sensor Networks with Distance and Angle Information". In: *10th International Conference on Information Processing in Sensor Networks (IPSN)*. IEEE. 2011, pp. 211–222.

[71] Prabal Dutta et al. "A Building Block Approach to Sensornet Systems". In: *In Proceedings of the Sixth ACM Conference on Embedded Networked Sensor Systems (SenSys)*. 2008.

[72] Aric A. Hagberg, Daniel A. Schult, and Pieter J. Swart. "Exploring Network Structure, Dynamics, and Function using NetworkX". In: *Proceedings of the 7th Python in Science Conference*. Ed. by Gaël Varoquaux, Travis Vaught, and Jarrod Millman. Pasadena, CA USA, 2008, pp. 11–15.

[73] Andrzej Forys et al. "Demo Abstract: Rapid Deployable System for Human Contact Network Research". In: *Proceeding of the 10th ACM Conference on Embedded Networked Sensor Systems (SenSys)*. 2012.

[74] Alessandro Montanari et al. "Measuring interaction proxemics with wearable light tags". In: *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2.1 (2018), pp. 1–30.

[75] *Motion Capture Systems*. URL: https://optitrack.com/.

[76] *WHO Save Lives: Clean Your Hands In The Context Of COVID-19*. https://www.who.int/infection-prevention/campaigns/clean-hands/WHO_HH-Community-Campaign_finalv3.pdf?ua=1. 2020.

[77] *Show Me the Science - Why Wash Your Hands?* https://www.cdc.gov/handwashing/why-handwashing.html. 2020.

[78] The Joint Commission. "MEASURING HAND HYGIENE ADHERENCE: OVERCOMING THE CHALLENGES". In: (2009).

[79] World Health Organization et al. *WHO guidelines on hand hygiene in health care (advanced draft): a summary: clean hands are safer hands*. Tech. rep. World Health Organization, 2005.

[80] M Musu et al. "Assessing hand hygiene compliance among healthcare workers in six Intensive Care Units". In: *Journal of preventive medicine and hygiene* 58.3 (2017), E231.

[81] World Health Organization et al. *WHO guidelines on hand hygiene in health care: first global patient safety challenge clean care is safer care*. World Health Organization, 2009.

[82] "A Survey of Handwashing Behavior (Trended)". In: *Prepared for: The American Microbiology Society & the American Cleaning Institute. Rochester, NY: Harris Interactive* (2010).

[83] Carl P Borchgrevink, JaeMin Cha, and SeungHyun Kim. "Hand washing practices in a college town environment". In: *Journal of environmental health* 75.8 (2013), pp. 18–25.

[84] Md Abu Sayeed Mondol and John A Stankovic. "Harmony: A hand wash monitoring and reminder system using smart watches". In: *proceedings of the 12th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services on 12th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*. 2015, pp. 11–20.

[85] Md Abu Sayeed Mondol and John A Stankovic. "HAWAD: Hand Washing Detection using Wrist Wearable Inertial Sensors". In: *2020 16th International Conference on Distributed Computing in Sensor Systems (DCOSS)*. IEEE. 2020, pp. 11–18.

[86] Xin Zhang et al. "Smart ring: a wearable device for hand hygiene compliance monitoring at the point-of-need". In: *Microsystem Technologies* 25.8 (2019), pp. 3105–3110.

[87] Valerie Galluzzi, Ted Herman, and Philip Polgreen. "Hand hygiene duration and technique recognition using wrist-worn sensors". In: *Proceedings of the 14th International Conference on Information Processing in Sensor Networks*. 2015, pp. 106–117.

[88] Hong Li et al. "Wristwash: towards automatic handwashing assessment using a wrist-worn device". In: *Proceedings of the 2018 ACM International Symposium on Wearable Computers*. 2018, pp. 132–139.

[89] *Dress Guidelines for Food Handlers*. http://cte.sfasu.edu/wp-content/uploads/2012/01/dressguidelinesforfoodhandlers.pdf. 2008.

[90] *What Food Handlers Can and Cannot Wear While on the Job*. https://foodhandlerscardhelp.com/what-food-handlers-can-wear-attire/. 2019.

[91] *Bare below the elbows guideline*. https://www.health.qld.gov.au/__data/assets/pdf_file/0026/680903/bare-below-the-elbows.pdf. 2017.

[92] A Hallas. *'Bare Below the Elbow' Supplementary Policy for Hand Hygiene*. https://www.tamesidehospital.nhs.uk/documents/BareBelowtheElbowsupplementpolicy.pdf. 2020.

[93] *BioVigil - Hand Hygiene Solutions*. https://biovigil.com/. 2020.

[94] *Hand Hygiene - SwipeSense*. https://www.swipesense.com/hand-hygiene. 2020.

[95] *AiRISTA Flow - RTLS Experts in Hand Hygiene*. https://www.airistaflow.com/solution/applicationhand-hygiene-compliance/. 2020.

[96] *Hand Hygiene Compliance Monitoring System for Hospitals*. https://www.ecolab.com/solutions/hand-hygiene-compliance-monitoring. 2020.

[97] *AeroScout Hand Hygiene Compliance Monitoring*. https://www.stanleyhealthcare.com/hospital-clinics/rtls/hand-hygiene-compliance-monitoring. 2020.

[98] *Electronic Hand Hygiene Monitoring*. https://www.vitalacy.com/hand-hygiene-monitoring-technology. 2020.

[99] *Centrak Hand Hygiene*. https://centrak.com/solutions/handhygiene-compliance/. 2020.

[100] Yen Lee Angela Kwok, Craig P Juergens, and Mary-Louise McLaws. "Automated hand hygiene auditing with and without an intervention". In: *American journal of infection control* 44.12 (2016), pp. 1475–1480.

[101] G Kinsella, AN Thomas, and RJ Taylor. "Electronic surveillance of wall-mounted soap and alcohol gel dispensers in an intensive care unit". In: *Journal of Hospital Infection* 66.1 (2007), pp. 34–39.

[102] *watchOS 7 adds significant personalization, health, and fitness features to Apple Watch*. https://www.apple.com/newsroom/2020/06/watchos-7-adds-significant-personalization-health-and-fitness-features-to-apple-watch/. 2020.

[103] *BAD001 SwipeSense Badge Test Report SwipeSense*. https://fccid.io/2AB5RBAD001/Test-Report/Test-Report-2324762. 2014.

[104] *T3s Staff Badge*. 2018. URL: https://www.stanleyhealthcare.com/sites/stanleyhealthcare.com/files/2018-10/AeroScout%5C%20T3s%5C%20Staff%5C%20Badge%5C%20Data%5C%20Sheet.pdf.

[105] Judith Dyson and Maurice Madeo. "Investigating the use of an electronic hand hygiene monitoring and prompt device: influence and acceptability". In: *Journal of infection prevention* 18.6 (2017), pp. 278–287.

[106] Saungi McCalla et al. "An automated hand hygiene compliance system is associated with improved monitoring of hand hygiene". In: *American journal of infection control* 45.5 (2017), pp. 492–497.

[107] Kate et al Brignall. "Quality of life and psychosocial aspects of cough". In: *Lung* (2008).

[108] National Center for Health Statistics. *2013 NAMCS summary tables*. https://www.cdc.gov/nchs/data/ahcd/namcs_summary/2013_namcs_web_tables.pdf.

[109] Giovanni A Fontana and John Widdicombe. "What is cough and what should be measured?" In: *Pulm. Pharmacol. Ther.* 20.4 (2007).

[110] Benjamin J et al. Cowling. "Comparative epidemiology of pandemic and seasonal influenza A in households". In: *New England journal of medicine* 362.23 (2010).

[111] Simon et al Cauchemez. "Household transmission of 2009 pandemic influenza A (H1N1) virus in the United States". In: *N Engl J Med* 2009.361 (2009).

[112] Allison E et al Aiello. "Design and methods of a social network isolation study for reducing respiratory infection transmission: The eX-FLU cluster randomized trial". In: *Epidemics* 15 (2016).

[113] Arnold S Monto et al. "Data resource profile: Household Influenza Vaccine Evaluation (HIVE) Study". In: *International journal of epidemiology* 48.4 (2019), 1040–1040g.

[114] Sophie et al Leconte. "Validated Methods of Cough Assessment: a Systematic Review of the Literature". In: *Respiration* (2010).

[115] Robert J et al Straka. "Patient self-reporting of compliance does not correspond with electronic monitoring: an evaluation using isosorbide dinitrate as a model drug". In: *Pharmacother* (1997).

[116] Jaclyn Smith and Ashley Woodcock. "New developments in the objective assessment of cough". In: *Lung* (2008).

[117] Michael A et al Coyle. "Evaluation of an ambulatory system for the quantification of cough frequency in patients with chronic obstructive pulmonary disease". In: *Cough* 1.1 (2005).

[118] Sergio et al Matos. "An automated system for 24-h monitoring of cough frequency: the leicester cough monitor". In: *IEEE Trans. Biomed. Eng* (2007).

[119] SS et al Birring. "The Leicester Cough Monitor: preliminary validation of an automated cough detection system in chronic cough". In: *Eur Respir J* (2008).

[120] Vinayak Swarnkar et al. "Neural network based algorithm for automatic identification of cough sounds". In: *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE. 2013, pp. 1764–1767.

[121] Digital Media Law Project. *Recording Phone Calls and Conversations*. https://www.dmlp.org/legal-guide/recording-phone-calls-and-conversations. 2020.

[122] Xiao et al Sun. "SymDetector: detecting sound-related respiratory symptoms using smartphones". In: *UbiComp 2015 Proceedings*. ACM. 2015.

[123] Eric C et al Larson. "Accurate and Privacy Preserving Cough Sensing Using a Low-Cost Microphone". In: *UbiComp'11*. ACM. 2011.

[124] Brian H Tracey, Germán Comina, et al. "Cough detection algorithm for monitoring patient recovery from pulmonary tuberculosis". In: *EMBC, 2011 Annual International Conference of the IEEE*. IEEE. 2011.

[125] Michael G et al Crooks. "Continuous Cough Monitoring Using Ambient Sound Recording During Convalescence from a COPD Exacerbation". In: *Lung* (2017).

[126] Nadia Yousaf et al. "Cough frequency in health and disease". In: *Eur Respir J* 41.1 (2013).

[127] C Freestone and R Eccles. "Assessment of the antitussive efficacy of codeine in cough associated with common cold". In: *J. Pharm. Pharmacol.* (1997).

[128] J Smith. "Cough: Assessment and equipment". In: *The Buyers Guide to Respiratory Care Products* (2008).

[129] Samantha J et al Barry. "The automatic recognition and counting of cough". In: *Cough* (2006).

[130] Mingyu et al You. "Cough detection by ensembling multiple frequency subband features". In: *Biomed Signal Process Control* 33 (2017).

[131] Eitan et al Kerem. "Effectiveness of PTC124 treatment of cystic fibrosis caused by nonsense mutations: a prospective phase II trial". In: *The Lancet* (2008).

[132] Joan M et al Daisey. "Indoor air quality, ventilation and health symptoms in schools: an analysis of existing information". In: *Indoor air* (2003).

[133] Angela Kelsall et al. "How to quantify coughing: correlations with quality of life in chronic cough". In: *Eur Respir J* 32.1 (2008).

[134] Cynthia L French et al. "Impact of Chronic Cough on Quality of Life". In: *Archives of Internal Medicine* 158.15 (1998).

[135] Jaclyn Smith and Ashley Woodcock. "Cough and its importance in COPD". In: *International journal of chronic obstructive pulmonary disease* 1.3 (2006).

[136] Anne B Chang and Peter G Gibson. "Relationship between cough, cough receptor sensitivity and asthma in children". In: *Pulm. Pharmacol. Ther.* (2002).

[137] AM Li, C Lex, et al. "Cough frequency in children with stable asthma: correlation with lung function, exhaled nitric oxide, and sputum eosinophil count". In: *Thorax* 58.11 (2003).

[138] Surinder S Birring, Sergio Matos, et al. "Cough frequency, cough sensitivity and health status in patients with chronic cough". In: *Respiratory medicine* 100.6 (2006).

[139] Gustavo et al Zayas. "Cough aerosol in healthy participants: fundamental knowledge to optimize droplet-spread infectious respiratory disease management". In: *BMC pulmonary medicine* 12.1 (2012).

[140] Benjamin J et al Cowling. "Aerosol transmission is an important mode of influenza A virus spread". In: *Nature communications* 4 (2013).

[141] William G et al Lindsley. "Viable influenza A virus in airborne particles expelled during coughs versus exhalations". In: *Influenza Other Respir Viruses* (2016).

[142] Wan Yang, Subbiah Elankumaran, and Linsey C Marr. "Concentrations and size distributions of airborne influenza A viruses measured indoors at a health centre, a day-care centre and on aeroplanes". In: *J. R. Soc. Interface* (2011).

[143] Thomas D Berry and Angela K Fournier. "Examining university students' sneezing and coughing etiquette". In: *American journal of infection control* 42.12 (2014).

[144] Robert G Loudon and Sharon K Spohn. "Cough frequency and infectivity in patients with pulmonary tuberculosis". In: *American Review of Respiratory Disease* 99.1 (1969).

[145] Richard D et al Turner. "Daily Cough Frequency In Tuberculosis Is Associated With Rates Of Household Infection". In: *ATS'16*. Am Thoracic Soc, 2016.

[146] J Korpáš, J Sadloňová, and M Vrabec. "Analysis of the cough sound: an overview". In: *Pulmonary pharmacology* 9.5 (1996).

[147] MJ et al Doherty. "The acoustic properties of capsaicin-induced cough in healthy subjects". In: *Eur Respir J* (1997).

[148] Paolo Michele Olia, Piersante Sestini, and Mario Vagliasindi. "Acoustic parameters of voluntary cough in healthy non-smoking subjects". In: *Respirology* (2000).

[149] Thomas et al Drugman. "Audio and Contact Microphones for Cough Detection." In: *INTERSPEECH*. 2012.

[150] Akira Murata, Yasuyuki Taniguchi, et al. "Discrimination of productive and non-productive cough by sound analysis". In: *Internal Medicine* 37.9 (1998).

[151] K et al. Kosasih. "High frequency analysis of cough sounds in pediatric patients with respiratory diseases". In: *EMBC'12*. IEEE. 2012.

[152] CW Thorpe, LJ Toop, and KP Dawson. "Towards a quantitative description of asthmatic cough sounds". In: *Eur Respir J* 5.6 (1992).

[153] *Factors Influencing Fundamental Frequency*. http://www.ncvs.org/ncvs/tutorials/voiceprod/tutorial/influence.html. 2020.

[154] Annemie Van Hirtum and Daniel Berckmans. "The Fundamental Frequency of Cough by Autocorrelation Analysis." In: *INTERSPEECH*. 2001, pp. 2435–2438.

[155] Tascam. *DR-05 Handheld Recorder*. http://tascam.com/product/dr-05/specifications/. 2017.

[156] *Microphone Frequency Response*. https://www.mediacollege.com/audio/microphones/frequency-response.html. 2020.

[157] *Shure SM57*. https://www.shure.eu/products/microphones/sm57. 2020.

[158] Jort F et al. Gemmeke. "Audio Set: An ontology and human-labeled dataset for audio events". In: *Proc. IEEE ICASSP 2017*. New Orleans, LA, 2017.

[159] J. Salamon, C. Jacoby, and J. P. Bello. "A Dataset and Taxonomy for Urban Sound Research". In: *(ACM-MM'14)*. Orlando, FL, USA, Nov. 2014.

[160] *You only cough when you're winning*. 1998. URL: http://news.bbc.co.uk/2/hi/health/166138.stm#:~:text=Coughing%5C%20can%5C%20be%5C%20as%5C%20loud,noisy%5C%20radio%5C%20or%5C%20underground%5C%20train.html.

[161] *What Noises Cause Hearing Loss?* https://www.cdc.gov/nceh/hearing_loss/what_noises_cause_hearing_loss.html. 2020.

[162] *ICD-PX370 Mono Digital Voice Recorder*. https://www.sony.com/electronics/voice-recorders/icd-px370/specifications. 2020.

[163] Giulia Regalia et al. "Multimodal wrist-worn devices for seizure detection and advancing research: focus on the Empatica wristbands". In: *Epilepsy research* 153 (2019), pp. 79–82.

[164] Shibo Zhang et al. "NeckSense: A Multi-Sensor Necklace for Detecting Eating Activities in Free-Living Conditions". In: *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4.2 (2020), pp. 1–26.

[165] Robert Triggs. *Fact check: Is smartphone battery capacity growing or staying the same?* 2018. URL: https://www.androidauthority.com/smartphone-battery-capacity-887305/.

[166] Christine Dierk. "Heirloom Wearables: A Hybrid Approach to the Design of Embodied Wearable Technologies". PhD thesis. EECS Department, University of California, Berkeley, July 2020. URL: http://www2.eecs.berkeley.edu/Pubs/TechRpts/2020/EECS-2020-138.html.